International Society of the Learning Sciences

# ICLS

## CHICAGO 2010

## Learning in the Disciplines

June 29 - July 2, 2010

9th International Conference of the Learning Sciences

hosted by

University of Illinois at Chicago
Learning Sciences Research Institute

# Conference Proceedings - Volume 1

## Full Papers

# LEARNING IN THE DISCIPLINES

## ICLS 2010 CONFERENCE PROCEEDINGS
## VOLUME 1- FULL PAPERS

9TH INTERNATIONAL CONFERENCE OF THE LEARNING SCIENCES

JUNE 28 - JULY 2, 2010, CHICAGO

UNIVERSITY OF ILLINOIS AT CHICAGO, ILLINOIS, USA

EDITORS:
Kimberly Gomez, Leilah Lyons, and Joshua Radinsky

Title: Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers

Editors: Kimberly Gomez, Leilah Lyons, and Joshua Radinsky

Cite as: Gomez, K., Lyons, L., & Radinsky, J. (Eds.) *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers*. International Society of the Learning Sciences: Chicago IL.



Published by:
International Society of the Learning Sciences (ISLS)

Proceedings Printed by:
LuLu
http://www.lulu.com

Proceedings Distributed by:
LuLu and Amazon

Book Cover Design:
Brenda Lopez Silva

## ICLS2010 Conference Sponsors

National Science Foundation

Learning Sciences Research Institute of the University of Illinois at Chicago

Northwestern University

Spencer Foundation

The Chicago Community Trust

MacArthur Foundation

National Geographic

The Research Journal for Engineering Education

Inquirium

# CONFERENCE ORGANIZATION

**Conference Co-Chairs**
Susan Goldman, *University of Illinois at Chicago, US*
James Pellegrino, *University of Illinois at Chicago, US*

**Business Manager**
Deana Donzal
*University of Illinois at Chicago, US*

**Communications Chair**
Leilah Lyons
*University of Illinois at Chicago, US*

**Program Co-Chairs**
Kimberly Gomez
*University of Pittsburgh, US*
Joshua Radinsky
*University of Illinois at Chicago, US*

**Workshops Co-Chairs**
Tom Moher
*University of Illinois at Chicago, US*
Eleni Kyza
*Cyprus University Of Technology, Cyprus*

**Special Sessions Chair**
Alison Castro-Superfine
*University of Illinois at Chicago, US*

**Doctoral Consortium Co-Chairs**
Cindy Hmelo-Silver
*Rutgers University, US*
Jerry Andriessen
*Utrecht University, Netherlands*

**Early Career Workshop Co-Chairs**
Susan Yoon
*University of Pennsylvania, US*
Paul Kirschner
*Open University, Netherlands*

**Conference Advisory Board Members**

| | |
|---|---|
| Kate Bielaczyc | *National Institute of Education, Singapore* |
| Ton DeJong | *Twente University, Netherlands* |
| Pierre Dillenbourg | *Swiss Federal Institute of Technology, Switzerland* |
| Daniel Edelson | *National Geographic Society, US* |
| Frank F. Fischer, | *University of Munich, Germany* |
| Kris Gutierrez | *University of Colorado at Boulder, US* |
| Janet Kolodner | *Georgia Tech University, US* |
| Joe Krajcik | *University of Michigan, US* |
| Eleni Kyza | *Cyprus University of Technology, Cyprus* |
| Marcia Linn | *University of California at Berkeley, US* |
| Claire OMalley, | *University of Nottingham, UK* |
| Nichole Pinkard, | *DePaul University, US* |
| Jim Slotta, | *Ontario Institute for Studies in Education* |
| Hans Spada, | *University of Freiburg, Germany* |
| Iris Tabak, | *Ben Gurion University of the Negev, Israel* |

**Program Committee Members**

| | |
|---|---|
| Dor Abrahamson | *University of California at Berkeley, US* |
| Angela Calabrese-Barton | *Michigan State University, US* |
| Ravit Duncan | *Rutgers University, US* |
| Cesar Cisneros Puebla | *Universidad Autónoma Metropolitana, Mexico* |
| Janice Gobert | *Worcester Polytechnic Institute, US* |
| Kim Lawless | *University of Illinois Chicago, US* |
| Paul Marshall | *The Open University, UK* |
| Naomi Miyake | *Chukyo University, Japan* |
| Brian K. Smith | *Rhode Island School of Design, US* |
| Reed Stevens | *Northwestern University, US* |

# CONTENTS

# Pictorial illustrations in intelligent tutoring systems: Do they distract or elicit interest and engagement?

Ulrike Magner, Rolf Schwonke, Alexander Renkl, Educational and Developmental Psychology, University of Freiburg, Germany
Email: ulrike.magner@psycholgie.uni-freiburg.de, rolf.schwonke@psychologie.uni-freiburg.de, renkl@psychologie.uni-freiburg.de
Vincent A.W.M.M. Aleven, Octav Popescu, Human-Computer Interaction Institute, Carnegie Mellon University, USA
Email: aleven@cs.cmu.edu, octav@cmu.edu

**Abstract:** Do pictorial illustrations distract from learning and, thus, decrease learning outcomes, as suggested by Cognitive Load Theory and Cognitive Theory of Multimedia Learning? Or can pictorial illustrations trigger interest and thereby enhance the willingness to learn as suggested by interest theories? Although these approaches seem to contradict each other, we assume that they are compatible: Pictorial illustrations may hamper short-term learning but raise interest and engagement so that in the medium run learning may be enhanced. In order to test our "integrative" hypothesis, we explored the potential of different types of pictorial illustrations to trigger situational interest in the context of geometry learning. Results showed an effect of pictorial illustrations on the interestingness of geometry problems. In addition, interest in further learning with the computer-based learning environment was higher with pictorial illustrations than without. On the other hand, interest in deepening geometry knowledge was reduced when illustrations were added.

## Introduction

Most (German) mathematical textbooks are full of pictorial illustrations. Does this make sense given that such pictorial illustrations may be "*seductive details*" that impede learning (Harp & Mayer, 1997, 1998)? Or do they actually enhance interest and, thereby, the willingness to engage in learning? Weidenmann (1991) suggests that pictorial illustrations have different supporting functions such as activation, construction, or focusing special aspects of the learning materials. In addition, Anglin, Towers, and Levie (1996) suggest in their review that pictorial illustrations–beside their compensatory, cognitive, and attentional functions–also have an affective function in learning.

Contrary to these assumptions on the beneficial functions of pictorial illustrations, *Cognitive Load Theory* (Sweller, 2005)–which assumes that many difficulties in learning are due to unnecessary (i.e., extraneous) working memory load–postulates a number of potential negative effects of pictorial illustrations (Sweller, van Merrienboër & Paas, 1998): Redundancy effect (i.e., when different sources such as pictures and text provide the same information and, thus, induce unnecessary processing demands), split-attention effect (i.e., extraneous load due to difficulties in mapping text and pictures), and extraneous load due to irrelevant illustrations that do not contain essential information. Similarly, *Cognitive Theory of Multimedia Learning* (Mayer, 2005) predicts that words, illustrations, or sounds which are not relevant to the central learning goals should be omitted (e.g., Harp & Mayer, 1997, 1998). In addition to the redundancy effect (Mayer, Heiser & Lonn, 2001; Moreno & Mayer, 2002) and split-attention effect (Mayer & Moreno, 1998; Moreno & Mayer, 1999; Moreno, Mayer, Spires, & Lester, 2001) Mayer and Moreno (2003) refer to the coherence effect (e.g., unnecessary load through interesting but extraneous material). In conclusion, learning materials should be as concise as possible (Mayer & Moreno, 2003). Both theories suggest eliminating any redundant or decorative elements. Accordingly, computer-based multimedia environments should not contain redundant or decorative pictorial illustrations.

In the last years, however, even cognitive load researchers increasingly call for the consideration of motivational and emotional aspects (e.g., Zander, Brünken, 2009; Paas, Tuovinen, van Merrienboër & Darabi, 2005; van Merrienboër & Ayres, 2005). Against this background, the construct of *interest* seems especially relevant (Tulis, 2009; Krapp, 1999, 2000; Krapp, Hidi & Renninger, 1992). It comprises a value-related and emotion-related component (Schiefele, Krapp, Wild & Winteler, 1992). Hidi (2000) distinguishes between two different, but linked types of interest, *individual* and *situational interest*. Individual interest is defined as a relatively stable predisposition, whereas situational interest arises as a reaction to environmental input, for example, visual and auditory (Krapp, 1999). Research has shown that (situational) interest eases comprehension (Hidi, 2001) and learning (Ainley, Hidi & Berndorff, 2002; Hidi, 2001; Schiefele, 1998). Explanations for such effects include deeper processing, and a higher degree of cognitive organization (Ryan, Connell & Plant, 1990). Interest is also positively related to persistence (Ainley, Hidi, 2002; Ainley, et. al., 2002), activation (Schiefele, 1990), and elaboration (Krapp, 1999; Schiefele, 1996, Ryan et al., 1990). Furthermore, it is essential for engagement in learning (Anderman, Noar, Zimmerman, & Donohew, 2004). Examples of factors that trigger and maintain situational interest are novelty, concreteness, and visual imagery (Hidi, 2001). Most results on situational and individual interest were found in the context of learning from text. *Concreteness, a personally relevant context, ease of comprehension*, and *unexpected information* (see Anderson, Shirey, Wilson & Fielding, 1987; Hidi & Baird, 1986, 1988) are, for example, factors that trigger and maintain situational interest in text learning. One way to adapt these aspects to computer-based multimedia environment is the use of concrete, meaningful and relevant context oriented pictorial illustrations.

Although the different theoretical perspectives on potential effects of pictorial illustrations seem to contradict each other, we assumed that they are compatible: Pictorial illustration may hinder short-term learning outcomes but can raise interest and engagement so that in the medium term learning is enhanced. As a first step in testing our "integrative" hypothesis, we conducted a study to explore the following aspects: First, we were interested whether pictorial illustrations can actually increase situational interest. Second, we wanted to identify sets of pictorial illustrations that do or that do not enhance situational interest in the domain of geometry. These pictorial illustrations will subsequently be used to test the integrative hypothesis in the context of learning with intelligent tutoring environments (here: Cognitive Tutors; Koedinger & Corbett, 2006). Up to now, most of the interest research has been conducted in the context of learning from text (Anderson, Shirey, Wilson & Fielding, 1987; Hidi & Baird, 1986, 1988). Our third goal was to test whether and how the four factors mentioned above would affect situational interest for the case of pictorial illustrations in the context of tutoring environments. A last goal was to identify potential relationships between picture-induced interest and the willingness to work with a picture-enriched computer-based learning environment as well as the willingness to deepen geometry knowledge.

## Method

### Participants and Design

Participants were 87 students (52 female, 35 male) from grade 8 of a German secondary school (age: $M = 13.9$ years; $SD = 0.6$). In a within-subjects design, students evaluated screenshots showing geometry problems with and without pictorial illustrations. In order to be able to evaluate a larger number of pictorial illustrations, we used a multi-matrix-design with eight questionnaires. Within each questionnaire one half of the screenshots corresponded to another half of the screenshots of another questionnaire.

More specifically, the main part of the questionnaire consisted of fifteen screenshots of a computer-based learning environment on the topic intersecting lines (mathematical principles: angle addition, vertical angles, linear pair, and complementary angle). Each screenshot showed a word problem and a corresponding line drawing. The problems varied in the number of sub-problems (one to three). Eleven of the fifteen drawings included pictorial illustrations that–from an instructional designer's point of view–were mainly decorative (i.e., without explicit instructional functions such as providing information or activating relevant schemas). The remaining four drawings did not include a pictorial illustration. However, each of these four drawings corresponded to one of the drawings that included a pictorial illustration.

### Materials

The screenshots were taken from a Cognitive Tutor Geometry (Koedinger & Corbett, 2006). The Cognitive Tutor is an Intelligent Tutoring System based on cognitive theory (Anderson, Corbett, Koedinger & Pelletier, 1995). Cognitive Tutors promote learning by tutored problem solving. They provide step-by-step feedback adapted to the actual knowledge level of students. Cognitive Tutors are successfully applied in a diverse set of domains such as mathematics, genetics, and computer programming (for an overview, see Anderson et. al., 1995; Koedinger & Corbett, 2006).

All pictorial illustrations were drawn from a pool of 44 illustrations. They showed real-life situations (e.g., sitting in front of computer, gables, and compasses). In the first part of the questionnaire, each page contained a screenshot of a geometry problem from the Cognitive Tutor lesson either with or without pictorial illustration (see Figures 1 and 2). Students had to rate the interestingness (adapted from Schiefele, 1990) of the screenshots on two different subscales (nine point Likert scales). The first subscale referred to the emotion-related component of interestingness, the second subscale to the value-related component (see Figures 1 and 2). The emotion-related component was assessed by the three items *Excitement*, *Entertainment*, and *Boredom,* the value-related component by the items *Usefulness*, *Worthless,* and *Unimportance* (similar items were used by Schiefele, 1990). Furthermore, for each screenshot the students rated (on nine point Likert scales) the previously mentioned four potential factors for interest (*concreteness, ease of comprehension*, *unexpected information*, and *personally relevant context*; see bottom part of figure 1 and 2).

Further questions (to be rated on five point Likert scales) referred to willingness for further learning with the computer-based learning environment (enriched with pictorial illustrations), willingness to deepen geometry knowledge, individual topic interest related to the content of the pictorial illustrations, to computer knowledge and mathematics knowledge and demographic characteristics.

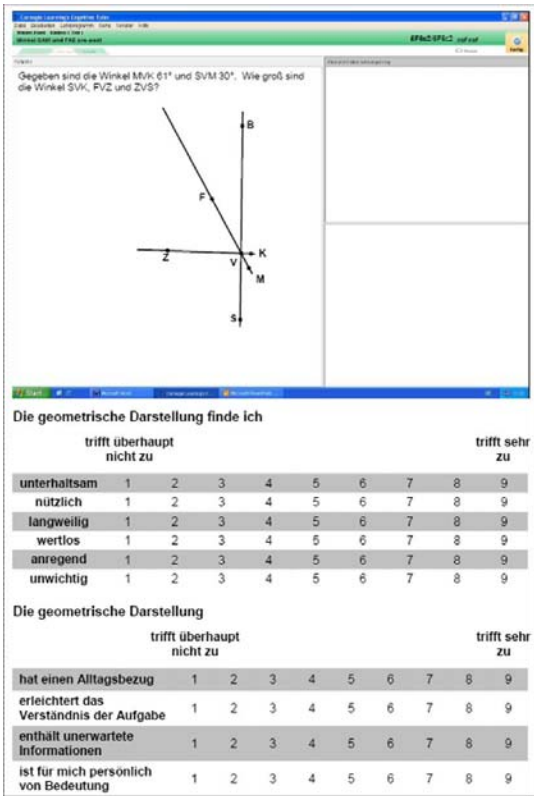



Figure 1. Page without pictorial illustration.

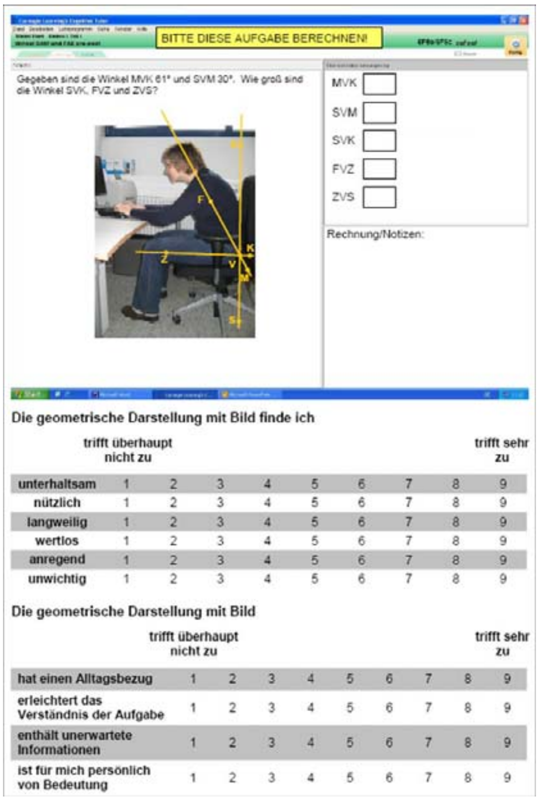



Figure 2. Page with pictorial illustration.

## Procedure

The study was performed in group sessions. The sessions lasted about 28 minutes (Range: 16 - 46 minutes). First, all students were introduced to the topic of intersecting lines by the experimenter. Then, they received an example of how to evaluate the screenshots. In order to provide a context as much similar as possible to actually working with the computer-based learning environment, the students then worked on four geometry problems (two with line drawings, two with line drawings enriched with pictorial illustrations). Afterwards the students rated the interestingness of the screenshots and they were asked to rate different potential reasons for their interest. Finally, all participants answered a set of additional questions (e.g., demographic characteristics).

## Results

Our first research question referred to a potential effect of pictorial illustrations on situational interest. A paired *t*-test (two-tailed) revealed a significant difference between screenshots (i.e., geometry problems) with and without pictorial illustrations, $t(85) = 4.02$, $p < .001$, $d = .55$. Figure 3 shows that this effect can largely be attributed to the emotion-related component of interest, $t(85) = 6.68$, $p < .001$, $d = .88$, as there were no differences in the value-related component, $t(85) = 0.85$, *n.s*, $d = .11$. Therefore, in all further analyses we concentrated on the emotion-related component (if not stated otherwise).

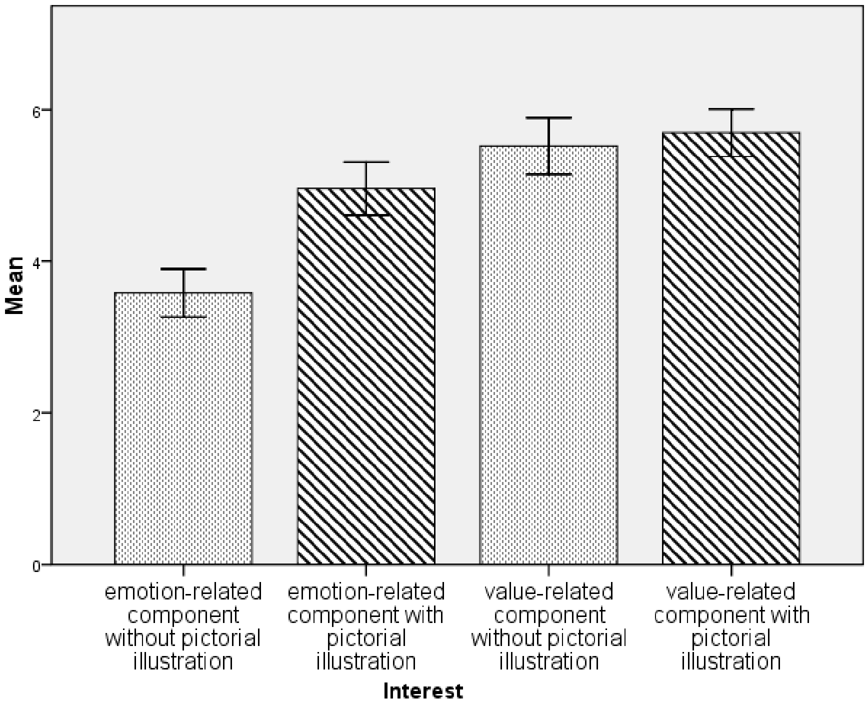



Figure 3. Effect of pictorial illustrations on interest (emotion-related and value-related)

Our second goal was to distinguish between the most interesting and the least interesting pictorial illustrations. We compared the effect sizes between the most interesting pictorial illustrations (upper quarter) and the least interesting pictorial illustrations (lower quarter). This comparison yielded an effect size of $d$ = .56 for the ninth most interesting pictorial illustration compared to the ninth least interesting one. Pictorial illustrations rated as "interesting" could be characterized as *dynamic* (e.g., showing activities such as sailing, volleyball, or riding). The most interesting picture, for example, showed a sailboat sailing close to the wind. Pictorial illustrations rated as "non-interesting" could be characterized as *static* (e.g., showing artifacts such as compasses, maps, or traffic signs). One of the least interesting pictures, for example, was a canvas. Therefore, it can be speculated that dynamic attributes are perceived as more entertaining, less boring and more exciting.

Our third goal was to assess whether and how factors that were found to increase interest in texts (i.e., concreteness, personally relevant context, ease of comprehension, and unexpected information) could be related to interest in tutoring environments that are enriched with pictorial illustrations. Table 1 shows that in the present study each of these four factors correlated significantly with both the emotion-related component and the value-related component of interest. This result supports the notion that these factors that are related to interest as induced by texts also have a psychological validity for interest as induced by pictorial illustrations (in tutoring environments).

Table 1: Correlations between reasons for interest and interest in drawings with and without pictorial illustrations ($N$ = 87)

| | Perceived concreteness | Perceived personally relevant context | Perceived ease of comprehension | Perceived unexpected information |
|---|---|---|---|---|
| *Emotion-related* | .46*** | .53*** | .55*** | .42*** |
| *Value-related* | .44*** | .45*** | .63*** | .29*** |
| | | | | |

***$p$ < .001.

Against the background of this strong pattern of correlations, we were interested whether and how these reasons would differ in situations (i.e., geometry problems) in which pictorial illustrations were available or not. Paired $t$-tests (two-tailed) showed that concreteness was rated higher in situations with pictorial illustrations than in situations without, $t(85) = 12.24$, $p < .001$, $d = -1.76$. Similarly, the perceived personally relevant context was rated higher in situations with pictorial illustrations $t(85) = -2.28$, $p < .05$, $d$= -0.21. There were, however, no significant differences for ratings of the ease of comprehension ($t(85) = 0.91$, *n.s.*, $d = .00$) or unexpected information, $t(85) = -1.98$, *n.s.*, $d = -0.20$.

In summary, this pattern of results indicates that decorative pictorial illustrations increased the perceived personal relevance and perceived concreteness of the learning material, but they did not increase the perceived ease of comprehension. This latter finding further indicates that the pictorial illustrations might not have activated relevant schemas as suggested, for example, by Schwartz and Collins (2008).

Finally, we were interested in whether pictorial illustrations can enhance further learning and engagement. For this purpose, we investigated the relationships between interest in situations with and without pictorial illustration and both the willingness to deepen geometry knowledge as well as in further learning with the illustration-enriched computer based learning environment. As can be seen in Table 2, the willingness to deepen geometry knowledge and the willingness for further learning with the computer-based environment were positively related to interest ratings (irrespective of the availability of pictorial illustration).

Table 2: Correlations between interest (with/without pictorial illustrations) and the willingness to learn with the computer-based learning environment and to deepen geometry knowledge (in Parentheses: partial correlations controlling for individual interest in geometry).

| | Willingness to learn with the computer-based learning environment | Willingness to deepen geometry knowledge |
|---|---|---|
| Interest (*n* = 84) | | |
| Without pictorial illustration | .24* (.15) | .45*** (.36)** |
| With pict illustration | .54*** (.47)*** | .22* (.04) |

*$p < .05$. **$p < .01$. ***$p < .001$

Comparisons of the correlation coefficients between mathematical problems with and without pictorial illustrations, however, showed that the correlations between interest and the willingness to learn with the computer-based learning environment was rated significantly higher when a pictorial illustration was available than when it was not, $t(84) = 2.58$, $p < .05$. This was not the case for the correlations of interest and the willingness to deepen geometry knowledge, $t(84) = 1.89$, *n.s*.

As it can be assumed that students who are generally interested in geometry might not need pictorial illustration to further increase their interest, we controlled the individual interest in learning geometry. The difference of the partial correlations (see Table 2, in parentheses) between interest and the willingness for further learning with the environment remained significant, $t(84) = 2.35$, $p < .05$. In addition, the partial correlation between interest and the willingness to deepen geometry knowledge became significant, $t(84) = 2.24$, $p < .05$.

Thus, interest as induced by pictorial illustrations seemed to be strongly positively related to students' perceptions of the learning environment more than to their willingness to deepen geometry knowledge. On the other hand, interest as induced by the mathematical problems per se (i.e., when no pictorial illustration was available) was strongly related to the willingness to deepen geometry knowledge; there was no such relationship for interest when a pictorial illustration was available. In other words, students who like pictorial illustrations might prefer to work with a computer-based learning environment but they are not necessarily interested in deepening geometry knowledge.

## Discussion

The main purpose of the present study was to identify pictorial illustrations with the potential to enhance interest. A secondary purpose was to test whether factors that were found to increase interest as suggested by text research (e.g., text attributes such as concreteness) might also enhance interest in the context of illustrated tutoring environments. A final purpose was to explore the relationships between interest and the students' willingness to engage in further learning.

The results showed that pictorial illustration increased the emotion-related component of interest, but not the value-related component. This suggests that the illustrations were mainly perceived as decorative but not as informative. Furthermore, our findings also suggest that interest-evoking attributes as identified in text research can be transferred to interest research with pictorial illustrations (in the context of tutoring environments), at least in part: The pictorial illustrations did not–according to the learners' perception–enhance comprehension or provide unexpected information. However, they were perceived as making the geometry problems more concrete and to provide a personally relevant context. These results are first indications that pictorial illustrations do not give a semantic context and, therefore, activate certain schemas which support learning (Schwartz & Collins, 2008). Further steps to investigate not only the perceived ease of comprehension, but real comprehension (in terms of conceptual and procedural knowledge) will be taken in a subsequent study.

Interestingly, the more the students perceived the pictorial illustrations as interesting the higher was their willingness to learn with a picture-enriched computer-based learning environment. Contrary, interest in pictorial illustrations was unrelated to the willingness to deepen geometry knowledge (especially when the individual geometry interest was held constant). The willingness to deepen geometry knowledge was rather related to interest in mathematical problems without pictorial illustrations. It can be speculated that students might have realized the "decorative" aspect of the pictorial illustrations, and thus, perceived them as irrelevant for further engagement in learning geometry. In conclusion, interesting pictures might primarily influence how much students engage in further working in a learning environment, but not necessarily how much they engage in concentrating on the learning content. In these respects, the present findings quite well reflect the contradictive positions between interest research and cognitive research. As suggested by interest research, pictorial illustrations can enhance interest and the willingness to learn. At the same time, however, students can perceive pictorial illustrations as irrelevant for deepening geometry knowledge, as suggested by the cognitive research.

A restriction of this study was that we did not measure learning outcomes. Moreover, the students did not work in a real computer-based learning environment (e.g., a Cognitive Tutor). Therefore, our conclusions are based on self-assessed measurements. An interesting question with respect to learning outcomes arises from the finding that interest elicited by decorative pictures seemed to foster the willingness for further working with a picture-enriched computer program but not to deepen geometry knowledge. If the learners would actually show certain persistence in working with an illustrated computer program, but without intentions to deepen the geometry knowledge, would this nevertheless deepen their understanding? A follow-up experiment (in preparation) will address this question. In addition, it will explicitly test our integrative hypothesis stating that pictorial illustration may hinder short-term learning but may raise interest and engagement so that in the medium term learning is enhanced. In order to test this assumption, we will compare three experimental groups (students from the 8th grade) working in a Cognitive Tutor Geometry (topic of intersecting lines): one group with interesting pictorial illustration, one with non-interesting pictorial illustrations (control of picture effects), and one group without any pictorial illustrations. While working with the computer-based learning environment the students will rate the interestingness of the pictorial illustrations after each geometry problem. Before and after the learning phase a geometry knowledge test (conceptual and procedural knowledge) will be administered. In addition, data on interest and mood experiences will be collected. During the two weeks between the immediate posttest and a delayed posttest the students will have the opportunity to deepen their geometry knowledge by studying a small geometry booklet (at home). The delayed posttest will be very similar to the first posttest but will additionally include problems from the booklet. We assume that in the immediate posttest learners will be best off without pictorial illustrations. For the delayed posttest, we expect, however, that interesting pictorial illustrations will lead to superior learning.

## References

Ainley, M., & Hidi, S. (2002). Dynamic measures for studying interest and learning. In P.R. Pintrich & M.L. Maehr (Eds.). *Advances in motivation and achievement: New directions in measures and methods.* (Vol 12, 43-76) Amsterdam: JAI.

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*(3), 545-561.

Anderman, E. M., Noar, S., Zimmerman, R., & Donohew, L. (2004). The need for sensation as a prerequisite for motivation to engage in academic tasks. In M.L. Maehr & P.R. Pintrich (Eds.)., *Advances in motivation and achievement, Volume 13: Motivating students, improving schools: The legacy of Carol Midgley*, (pp. 1-26). Greenwich, CT: JAI Press.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*, 167–207.

Anderson, R. C., Shirey, L. L., Wilson, P. T., & Fielding, L. G. (1987). Interestingness of children's reading material. In Snow, R. E., and Farr, M. J. (eds.), *Aptitude, Learning and Instruction*: *Vol. III, Cognitive and Affective Process Analyses*, Erlbaum, Hillsdale, NJ.

Anglin, G. J., Towers, R. L., & Levie, W. H. (1996). Visual message design and learning: The role of static and dynamic illustrations. In D. H. Jonassen (Ed.), *Educational communications and technology.* (pp.755-794).New York: Simon & Schuster Macmillian.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*(1), 92-102.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414-434.

Hidi, S. (2000). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and Extrinsic Motivation* (pp. 309-339).

Hidi, S. (2001). Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review, 13*(3), 191-209.

Hidi, S., & Baird, W. (1986). Interestingness—A neglected variable in discourse processing. *Cogn. Sci.* 10: 179–194.
Hidi, S., & Baird, W. (1988). Strategies for increasing text-based interest and students' recall of expository texts. *Reading Res. Q.* 23: 465–483.
Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp.61-77). New York, NY: Cambridge University Press.
Krapp, A. (1999). Interest, motivation and learning: An educational-psychological perspective. *European Journal of Psychology of Education, 14*, 23-40.
Krapp, A. (2000). Interest and human development during adolescence: An educational psychological approach. In J. Heckhausen (Ed.), *Motivational Psychology of Human Development* (pp. 109-128). Amsterdam: Elsevier.
Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning and development. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The Role of Interest in Learning and Development* (pp. 3-25). Hillsdale, NJ: Erlbaum.
Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York, NY: Cambridge University Press.
Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 390-397.
Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43-52.
Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.
Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*, 156-163.
Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based multimedia learning: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*, 177-214.
Paas, F., Tuovinen, J. E., Van Merrienboër, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learning involvement in instruction. *Educational Technology Research and Development, 53*(3), 25-34.
Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in nondirected text learning. *Learning and Individual Differences, 2*, 1-17.
Schiefele, U. (1990). Thematisches Interesse, Variablen des Leseprozesses und Textverstehen. [topic interest, variables of reading processes and text comprehension]. *Zeitschrift für experimentelle und angewandte Psychologie, XXXVII(2)*[Journal of experimental and applied psychology], 304-332.
Schiefele, U. (1996). Topic interest, text representation, and quality of experience. *Contemporary Educational Psychology, 21*, 3-18.
Schiefele, U. (1998). Individual interest and learning, what we know and what we don't know. In Hoffman, L., Krapp, A., Renninger, K., and Baumert, J. (eds.), *Interest and Learning: Proceedings of the Seeon Conference on Interest and Gender*, IPN, Kiel, Germany, pp. 91–104.
Schiefele, U., Krapp, A., Wild, K.-P., & Winteler, A. (1992). Eine neue Version des "Fragebogens zum Studieninteresse" (FSI) [New version of "study interest questionnaire"]. Gelbe Reihe-Arbeiten zur Empirischen Pädagogik und Pädagogischen Psychologie (21) [Yellow Issue-Works of empirical education and educational psychology], 1-26.
Schwartz, N. H., Collins, C. (2008). Arguing a position from text: The influence of graphic themes on schema activation. In J. Zumbach, N. Schwartz, T. Seufert, & L. Kester (Eds.), Beyond knowledge: The legacy of competence. Meaningful computer-based learning environments (pp. 117-119). Berlin: Springer.
Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 19-30).
Sweller, J., van Merrienboër, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review 10* (3), 251-296.
Tulis, M. (2009). Interest, enjoyment and pride after success and failure in mathematics. Fostering Communities of Learners. Paper presented at the 13th Biennial Conference for Research on Learning and Instruction. Amsterdam (NL), August 2009.
Van Merrienboër, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development, 53*(3), 5-13.
Weidenmann, B. (1991). *Lernen mit Bildmedien.* [Learning with visual media]. Weinheim, Basel: Beltz Verlag.

Zander, S., & Brünken, R. (2009). Do intrinsic and extrinsic aspect of motivation affect cognitive load in different ways? Fostering Communities of Learners. 13th Biennial Conference for Research on Learning and Instruction, 477. Paper presented at the 13th Biennial Conference for Research on Learning and Instruction. Amsterdam (NL), August 2009.

## Acknowledgments

In particular we thank Hans-Dieter Baumgärtner for his practical support and making this study possible.

# Social Software and Knowledge Building: Supporting Co-Evolution of Individual and Collective Knowledge

Joachim Kimmerle, University of Tuebingen, Adenauer-Str. 40, 72072 Tuebingen, j.kimmerle@iwm-kmrc.de
Ulrike Cress, Knowledge Media Research Center, Adenauer-Str. 40, 72072 Tuebingen, u.cress@iwm-kmrc.de
Christoph Held, Knowledge Media Research Center, Adenauer-Str. 40, 72072 Tuebingen, c.held@iwm-kmrc.de
Johannes Moskaliuk, University of Tuebingen, Adenauer-Str. 40, 72072 Tuebingen, j.moskaliuk@iwm-kmrc.de

**Abstract:** This paper presents a framework model that defines learning and knowledge building as a co-evolution of cognitive and social systems. This model brings together Scardamalia and Bereiter's theory of knowledge building and Nonaka's knowledge creation theory. We demonstrate how learning and knowledge building may occur when people interact with each other, using shared digital artifacts such as tag clouds (that result from social-tagging activities) or wikis. For both technologies, we provide illustrating data from two pilot studies. As an example, we refer to the learning processes that take place while searching for information in tag clouds. In addition, we illustrate processes of knowledge building by referring to users working on a wiki. In conclusion, the differences and similarities between these technologies are assessed, regarding their potential for knowledge building.

## Introduction

Recent developments of software technology have made new tools available, which are of great importance for computer-supported collaborative learning and knowledge building. They provide new opportunities for learning and knowledge building, because they are capable of facilitating the interplay between individual and collective processes. These technologies are associated with social-software systems. *Social software* is a term for software systems that support human communication, collaboration, and interaction in large communities (Kolbitsch & Maurer, 2006). They facilitate the establishment and maintenance of self-organizing communities and social networks (Köhler & Fuchs-Kittowski, 2005; Wasko & Faraj, 2005). Social software is mainly associated with Internet communities, but may also be applied in educational contexts (Notari, 2006; Wang & Turner, 2005). We believe that social software has a great potential in the field of learning and knowledge building.

For a long time, psychological and educational research have examined individual learning and collaborative knowledge building as two separate things (for an early systematic model of collaborative knowledge building cf. Stahl, 2000). We propose that software tools are now able to support "interdigitation" and, so to speak, a "merger" of individual and collective knowledge processes. This new development of software technology also calls for a new theoretical framework, in order to focus on the tight conjunction between individual learning and collective knowledge building, in the sense that the knowledge of individuals and of a community may cross-fertilize and mutually support the development of each other. The aim of this paper is to provide such a theoretical framework. For this purpose, we will present and bring together two theoretical approaches.

The first approach is Scardamalia and Bereiter's (1991) theory of knowledge building. It deals with the mechanism through which a community of learners will manage to develop knowledge jointly. This approach was already based on the use of computer technology, and it emphasized the impact of epistemic artifacts on knowledge building. So it appears to be quite a suitable theory to explain how social software can support the development of collective knowledge. The second approach is Nonaka and Takeuchi's (1995) knowledge creation theory. Its focus is on the building and transfer of tacit knowledge. This approach is very popular in literature on knowledge management, but has hardly received attention in educational research and the learning sciences. Although Nonaka's model does not explicitly mention computer support and the creation of artifacts, we believe that the major underlying ideas of this model are very useful to explain some details of knowledge building with social software.

We will finally present our own co-evolution model of cognitive and social systems. The model takes into account social processes that are facilitated by a collaboratively developed technical artifact and, at the same time, cognitive processes of the individual users. Our model demonstrates how these social and cognitive processes will mutually influence each other. It is a theoretical framework which integrates many of the processes that have been described by the theory of knowledge building and the theory of knowledge creation. After a more general description of what we mean by "co-evolution" of cognitive and social systems, we will provide two examples of social-software technologies, and present some empirical data to demonstrate how these technologies can support this process of co-evolution. For this purpose, we will describe the use of social-tagging systems and the application of wikis. Each of these technologies creates a unique kind of artifact, which supports, in its own specific way, those processes of knowledge building which we have described. In conclusion, we will assess the differences and similarities between these technologies, regarding their potential for knowledge building.

## Theories of Knowledge Building and Knowledge Creation

Scardamalia and Bereiter have proposed a theory of knowledge building (Bereiter & Scardamalia, 1996; Scardamalia & Bereiter, 1996, 1999, 2003). Knowledge building is defined as a socio-cultural process, which takes place in a community and aims "at producing something of value to the community – theories, explanations, problem formulations, interpretations, and so on, which become public property that is helpful in understanding the world" (Scardamalia & Bereiter, 1999, p. 276). Scardamalia and Bereiter consider knowledge building as a collective creation of public knowledge. They compare learning in classrooms to learning in knowledge-building communities, such as research laboratories (Bereiter, 2002), and conclude that knowledge building should be rendered possible in all communities. Educational software, such as CSILE (Computer-Supported Intentional Learning Environments) or Knowledge Forum, can support knowledge building. These environments provide shared databases as collaborative design spaces. Here, all participants may contribute their own theories, models, examples, visualizations, notes, and other epistemic artifacts. The design space supports mutual citing and referencing, in order to initiate a dynamic and self-organized process in which ideas are formulated, discussed, revised, or rejected. The design space visualizes this collective improvement of ideas. Accordingly, there is a set of requirements that have to be fulfilled to make successful knowledge building possible (Hewitt & Scardamalia, 1998; Scardamalia, 2002). All members of a community should contribute to the advancement of knowledge, and learners in a community should be concerned with authentic real-life problems. Advancing knowledge should be regarded as an improvement of ideas, not a search for a perfect or true solution. Learning is regarded as a discourse-oriented process, in the sense of common problem solving.

Nonaka's knowledge-creating theory (Nonaka, 1991, 1994; Nonaka & Takeuchi, 1995) is also concerned with innovation. This theory emerged from research on knowledge management, and it regards knowledge dissemination as one of the aims of any knowledge organization. But the authors assume that within an organization only little knowledge is available in the form of explicit knowledge. Most knowledge is contained in the experience of individuals, and as such it is *tacit knowledge* (Polanyi, 1966). Tacit knowledge can hardly be expressed verbally, so this type of knowledge is difficult to transfer to others. But effective knowledge creation requires such a transfer. In order to describe how a transfer of tacit knowledge takes place, the authors refer to four processes, which build on each other in a dynamic way: socialization, externalization, combination, and internalization. *Socialization* is a process in which the experience of an individual is shared. Tacit knowledge can only be communicated from one person to the other through direct experience. Experience may be transferred by observation and imitation, and, as a result of this transfer, the observer acquires new knowledge, but this remains tacit. So a process of *externalization* is necessary. It consists of the articulation of tacit knowledge and the transfer from tacit to explicit knowledge. At this stage, new concepts are formulated using tacit knowledge. When knowledge has been made explicit, then it can be combined, in order to develop new knowledge within an organization. So, *combination* refers to connecting and giving a new structure to explicit knowledge. Such combination is supported by interpersonal sharing of knowledge. Explicit knowledge will then, once more, become tacit knowledge through a process of *internalization*. This conversion is an individual process, in which explicit knowledge becomes part of that person's personal knowledge. Internalization will occur through learning by doing. Nonaka has formulated the idea of a *knowledge spiral* in which socialization, externalization, combination, and internalization succeed each other. The spiral allows the inter-individual transfer of knowledge and, at the same time, the creation of new knowledge, through a combination of externalized knowledge which was previously tacit.

## Co-Evolution Model of Cognitive and Social Systems

The authors of this paper have introduced a theoretical model of individual learning and collaborative knowledge building (Cress & Kimmerle, 2007, 2008; Kimmerle, Moskaliuk, & Cress, 2009; Moskaliuk, Kimmerle, & Cress, 2008). In this model, collaborative knowledge building is described as an interplay between cognitive systems and a social system. Just like Scardamalia and Bereiter, we think that knowledge building is a central issue in modern knowledge societies. Consequently, our model is basically predicated on the ideas of Scardamalia and Bereiter. But our considerations go beyond those of these authors, as we focus equally on individual learning processes and on the collective development of knowledge. And as far as social software is concerned, epistemic artifacts are not only considered as a means to an end, but also as an end in itself, because we are convinced that collective knowledge manifests itself in shared digital artifacts. The model borrows, so to speak, from systemic and from cognitive approaches. Its main assumption is that knowledge building can only be understood if the interplay between individuals and the collective is taken into account. From a systems-theoretical point of view – in the sense of Luhmann (1995) – the cognitive systems of individuals are different from a social system, which is (in our theory) represented by a shared digital artifact (Cress & Kimmerle, 2008).

Cognitive systems and social systems have different kinds of operations. Due to their different modes of operation, both systems cannot simply merge. But one system can affect the other one in its development by irritating it. Each system can provide its own complexity for the development of the other.

The co-evolution model of cognitive and social systems (Cress & Kimmerle, 2008) is based on this systemic view, and it describes learning processes and knowledge-building processes as reactions of a system to irritations. Irritations are interpreted, in the sense of Piaget (1977), as cognitive conflicts, and the assumption is that cognitive systems will develop when people solve such cognitive conflicts. A cognitive conflict exists when people's prior knowledge and information which they receive from their environment are somewhat incongruent. Cognitive conflicts can be solved by processes of equilibration. There are two types of such equilibration processes: people will either assimilate information, i.e. they simply add new information to their prior knowledge; or they will accommodate their prior knowledge to new information. In either case, people will somehow have to internalize information from their environment. In this way, their cognitive system becomes more complex. This development of a cognitive system refers to what is traditionally called *learning*. These two processes represent two types of internalization from the social system, i.e. both *assimilation* and *accommodation* support the development of cognitive systems. The model states that processes which take place in a cognitive system by means of internalization will take place analogously in a social system by means of externalization. Through externalization, a cognitive system can bring that person's own individual knowledge into the shared artifact. People do not only internalize information from their environment into their cognitive systems, but also externalize their knowledge. This means, just as individuals can learn by internalizing new information, social systems can also learn by incorporating information, and social systems can develop new knowledge by assimilation or accommodation respectively. They can develop by just adding new content (assimilation) or by changing their own structure (accommodation). We propose that emergent effects usually occur through such accommodations of artifacts. This external accommodation leads to a higher complexity of the shared digital artifact and, accordingly, to new equilibration processes in other people's cognitive systems. Thus, the users' knowledge supplies the artifact's content, and the artifact itself provides its users with new information and releases new cognitive conflicts. What is important here is that the processes of internalization and externalization, or of individual learning and knowledge building respectively, are not independent from each other. It is always a matter of internalization *and* externalization, i.e. of continuous exchange processes between cognitive systems and the social system. In this way, there exists a mutual development of cognitive and social systems, and this is what we refer to as co-evolution (cf. Figure 1).

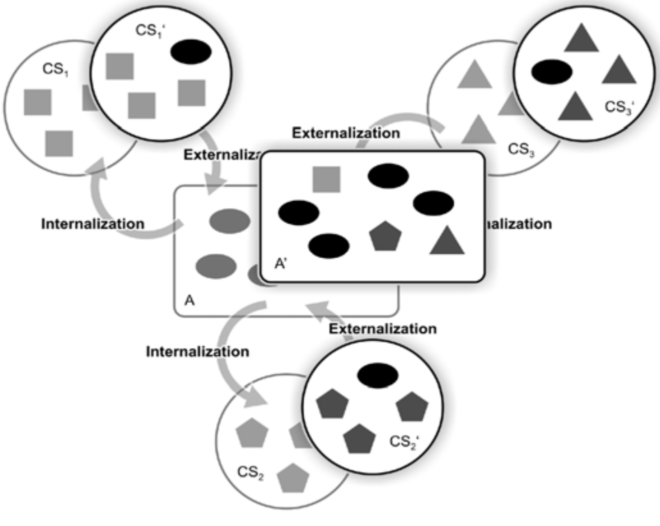



Figure 1. Co-evolution of cognitive systems ('CS') and a social system, i.e. an artifact ('A').

We can now apply the four processes, as they were described by Nonaka, to this model of knowledge building with shared artifacts. For this purpose, we translate the processes of making tacit knowledge explicit, and vice versa, into processes of externalization and internalization. This will help us to understand exchange processes between cognitive systems and the social system in more detail. Externalization can then be described by those processes which Nonaka has labeled *externalization* and *combination*. When people introduce their own knowledge into a shared digital artifact, they will have to articulate their knowledge, i.e. they need to couch their cognitive concepts in words. For this purpose, they have to translate their knowledge into a form that is generally understandable and intelligible to all. People have to consider the information which is already available in an artifact in order to integrate their own thoughts adequately. Internalization combines the two processes which Nonaka calls *internalization* and *socialization*. When people incorporate information from a shared digital artifact, they "dig around" for information by browsing the artifact. In this way, they gather and collect relevant information, which they have to transfer into their own cognitive structures. In the following we will explicate these processes by referring to two concrete examples of shared digital artifacts. The first example – social tagging – considers the process of internalization, the second example – wikis – will take a closer look at externalization.

## Examples of Social-Software Supported Knowledge Building

We present our notions by providing two examples that use new software technology. The first one will illustrate the interplay between individual and collaborative processes by analyzing *social-tagging activities* and the processes which may be supported by social tagging. Then, we will describe how *wikis* can contribute to knowledge building. For both technologies we point out their potential as learning environments, and demonstrate the processes of internalization and externalization. For this purpose, we provide some empirical data from two pilot studies: processes of internalization (individual learning) will be illustrated by a reference to searching for information in tag clouds; processes of externalization (knowledge building) will be demonstrated by quoting contributions of people who worked on a wiki.

### Social Tagging

Social tagging has become a standard feature of many social-software tools, and it has potential to play a role in learning scenarios (Marlow, Naaman, Boyd, & Davis, 2006). The term tagging describes the annotation of digital resources with keywords (Golder & Huberman, 2006). These resources are primarily objects in virtual environments, like the Internet or intranets, and they may include web sites, photos, videos, research papers, or other pieces of digital information. The keywords (or tags) may be chosen by its users, and they represent the users' associations concerning the objects. Tagging is used on platforms which offer file sharing within a community, or which support users in organizing their own stored resources. The labeling of items with tags may facilitate search processes and the retrieval of information on these platforms, not only for individual users, but also for the benefit of others. The term *social* tagging refers to this social context in which tags are used, and in which all members of a community may benefit from them. Most of the social-tagging systems allow each user to tag all available resources individually and independently. In this way, all tags in a community that concern a particular object may be aggregated, and a large set of metadata for that resource may come together. This accumulated collection of tags (or metadata) represents the users' connections, concepts, or categorizations for items, and builds a jointly created artifact of keywords that refer to a specific resource. The World Wide Web provides loads of information and many opportunities for acquiring knowledge. This can take place as an individual and independent learning process, beyond formal educational settings. Users may browse through the information space of the Internet, and access information on virtually any topic of interest. Limitations lie mainly in the overwhelming amount and inconsistent quality of that information. Social-tagging systems can help to overcome these problems and provide a tool for structuring and filtering information.

When dealing with social-tagging systems, individuals have the opportunity to externalize their own knowledge and, at the same time, they also may internalize information. In the process of externalization, users add tags to specific resources and describe the object with their own set of keywords. For this activity of creating tags, users have to articulate their own cognitive concepts of an item, and transform them into keywords. When people add tags, they externalize their knowledge about a resource. They focus on the essential concepts of an item, and may, consequently, elaborate on the resource more thoroughly. The additional cognitive effort which is involved here will induce processing of information in such a way that individual learning takes place. The creation of tags and externalization of cognitive concepts can lead to knowledge building (Budiu, Pirolli, & Hong, 2009): the artifact of metadata develops in a step-by-step incremental process of individual tagging. All individual tags of a community are summed up in one accumulated artifact of keywords. A single user has only a marginal influence on the whole product. The artifact depends on the tagging activities of numerous people, and tag clouds will only emerge from a large quantity of added keywords. As the artifact grows, it represents the most important related concepts and connections that concern a resource or a keyword.

In the process of internalization, people integrate new information into their cognitive systems. In a social-tagging environment, users will browse a web site in order to explore relevant resources. For this purpose they may use tags as an orientation and searching device. The jointly created artifacts of metadata, typically visualized as tag clouds, help them to navigate through the information space and find relevant information related to the tags. In this process of browsing, the users of the system will become aware of the keywords that were annotated by other users. In this way, the users incorporate information about the concepts and categorizations of an object, and learn how other users have classified that resource. Moreover, the artifact of tags may show interconnections between resources and concepts that may previously have been unknown to the individual user. The information which is represented in the metadata of tags can lead to knowledge acquisition and a change in the cognitive structures of individuals. People can be made aware of new information and incorporate the community's concepts and categories (Fu, 2008). These concepts and connections may differ from individual cognitive structures, and if they do, individual knowledge may increase and new understanding may develop. This process of learning may be defined as assimilation or accommodation of knowledge. Assimilation describes the process of merely adding new pieces of information to previously existing knowledge. This happens when people acquire additional facts about a subject, but do not significantly change the underlying structure of the cognitive concepts involved. For instance, when users learn new tags and related concepts which are in accordance with their prior knowledge about a certain resource, this will extend their factual knowledge, but they will not form a new concept of this subject. The following example will illustrate this process. The data are taken from a pilot study with university students. Participants had to find out what "EMDR (eye movement desensitization and reprocessing)" means, by searching and browsing for information only with the help of tag clouds and, at the same time, they had to think aloud. The following transcript from the thinking-aloud audio protocols (translated from German) is an example of assimilation:

> *"Yes, this is a form of psychotherapy, I know that, that's why I click on 'therapy' […] yes, exactly, and there is 'treatment', there I will get it in more detail […] okay, now, here is 'trauma', it has something to do with that […] and with 'anxiety' […] and it could also be about 'stress' […]".*

This participant expanded his knowledge, but he relied very much on his prior knowledge (that EMDR is a form of psychotherapy). So he specified his knowledge: later, he knew what this therapy is used for (the treatment of post-traumatic stress disorder). Previously existing knowledge was supplemented, but the knowledge structure did not have to be changed. In contrast to this process of assimilation, accommodation takes place when a qualitatively new understanding of a subject develops and prior knowledge is transformed. If other people use very different tags from those that this particular user would have applied, then it is obvious that specific resources or tags are related to very different concepts. So users may learn that their associations on a subject or topic were incorrect, and may change their cognitive concepts accordingly. Another sample from the thinking-aloud protocols of the pilot study will illustrate the accommodation process:

> *"I think it is about an eye movement disorder […] okay, here is 'movement', I click on that […] now there is 'desensitization' […] now 'reprocessing' remains as the biggest tag […] and, yes, 'treatment' […] and now it's about 'behavior' and eh? Hm, 'trauma'? Now I'm confused, somehow I can't see any connection. But I would click on that, on 'trauma', yes. It somehow stands out from the crowd […]".*

Even though this participant did not find out (in the given time) that EMDR is a form of psychotherapy, she did understand that her prior knowledge was not adequate, which is an important prerequisite for accommodation. Another participant had a similar experience, but her gain of knowledge went even further:

> *"[…] 'anxiety'? I'm irritated now […] I don't understand […] I just see that it has something to do with 'psychiatry' and with 'clinical' […] I see, it seems it is about psychology and psychotherapy, and, although I first thought it was an eye disease […]".*

The process of co-evolution of cognitive and social systems takes place when individuals use tags for browsing and navigating the Internet, and stumble across relevant information and resources. These cognitive processes of internalization and the retrieval of new information may prompt users to tag the discovered new resources themselves. In this way, the individuals' concepts and categorizations are externalized and incorporated into the artifact of metadata, and the whole system evolves (Fu, 2008). All members in a social-tagging community offer and obtain knowledge, and a continuous process of advancing knowledge takes place. The community constantly adds new resources and new tags to the artifact, in this way developing new information, new interconnections, and new ideas. Anyone can contribute and participate in this process of improvement, which takes up the viewpoints of many different people; the collective as a whole is responsible for the advancement of knowledge.

## Wikis

A wiki is a web page that allows users to change its content online (Leuf & Cunningham, 2001). It is easily accessible and can simply be used by everyone (Désilets, Paquet, & Vinson, 2005). In a wiki, people may easily revise all parts of the text, add, change, or delete anything at their discretion (Raitman, Augar, & Zhou, 2005). In that way, people can form communities that work collaboratively on a certain topic and create new content (Köhler & Fuchs-Kittowski, 2005; Moskaliuk & Kimmerle, 2009). Wikis may be used in educational contexts (Notari, 2006; Wang & Turner, 2005) – schools, universities, but also informal learning settings. A wiki lends itself, for example, to collective work on a scientific topic in school or a university class, particularly if people tend to have controversial opinions on that topic. In a wiki, those involved can introduce their own points of view on equal terms. Participants may express opposing opinions, addressing each other, and they incorporate their own perspective into a coherent text. But such processes of knowledge acquisition and knowledge exchange will not only take place in formal settings but also in informal contexts, for example, when people try to deal with and develop a topic through the Internet. This is a way for them to increase their own understanding, and they are able to acquire new knowledge. In addition to their individual learning, they also develop some collective knowledge.

Individuals may acquire new knowledge when they internalize information from the wiki. In order to incorporate information from the artifact, they will have to start off by browsing for information in a more or less target-oriented manner. When they have found some relevant information, they will have to transfer it into their own cognitive system. They will have to treat this information in some way or other to fit it into their own existing knowledge. So this internalization process allows people to increase their individual knowledge and gain new insights. This type of individual learning may, again, either take place in the form of assimilation or accommodation. When people assimilate information, they simply acquire knowledge without developing a different quality of understanding. Accommodation will occur when they incorporate new information in such a way that it modifies their prior knowledge. In an accommodation process, people transform knowledge with the aim of better understanding new information. In addition to this individual learning, as a result of internalization, there is also a supplementary form of knowledge generation: people will not only internalize information as it is, but also develop knowledge that is completely new, knowledge that was originally neither part of the digital artifact nor a component of their individual knowledge. This supplementary knowledge may develop when people have internalized new information, which will then, in turn, interact with their prior knowledge. This is the way in which a process of emergence is enabled. In such a situation, collaboration is a prerequisite for emergent knowledge. People on their own will not be able to develop this kind of knowledge; they need stimulation by other people. So knowledge processes enabled by a wiki are not simple knowledge-sharing processes, but collaborative processes of knowledge development.

When people want to contribute to a wiki, they will have to externalize certain aspects of their own knowledge. People who introduce their own knowledge to a shared digital artifact will have to put some effort into articulating that knowledge. That means they have to transform cognitive concepts into written language. Ideally, people will express their thoughts in a way that can easily be understood by others. They need to consider pre-existing information in the artifact to make sure that they integrate their own thoughts adequately. When people contribute to a wiki article, this will not only support the development of the wiki itself, but also its contributors' individual learning: people who externalize their knowledge will have to elaborate on it. This cognitive effort helps people to process their own knowledge more deeply, and this, in turn, will broaden their understanding (cf. the *self-explanation effect*; Chi, Bassok, Lewis, Reimann, & Glaser, 1989). When individuals have externalized their own knowledge, the resulting information continues to exist independently from its contributors. This information may, in turn, be picked up by other users, who may develop it further.

When people introduce knowledge into a wiki, there are two options for incorporating this information into the artifact. The information is either assimilated or accommodated. It is a matter of assimilation when information is simply attached to a text that already exists. In this case, new information is not connected to earlier information. Assimilation means that the previous arrangement of the artifact remains unchanged. An example will illustrate this process. The data are derived from another pilot study in which participants had to deal with a wiki text about causes of schizophrenia. The original wiki text was rather one-sided – only presenting biological causes, such as inheritance – and some participants only added information to the text that argued along the same lines, even though they had access to a variety of information sources; for example, the original text stated that an inherited disposition caused schizophrenia, and a participant wrote:

> *"This is supported by a study of the University of Göttingen about the genetic influence on suffering from schizophrenia. The results show that the probability of being afflicted with schizophrenia is increased when someone is consanguineous with a schizophrenic person. The probability of being afflicted with schizophrenia is 9.35 % for children. Another study with monozygotic twins shows that…".*

This contribution may be considered an essential addition, because it is important to provide empirical data for claims and statements. But it contains only supplementary information; it does not provide an alternative point of view. In a well-organized wiki, however, the process goes beyond such a simple assimilation of information. An efficient wiki community will endeavor to interpolate new information evenly. A wiki reflects accommodation if newly-introduced information is not just added to some existing text, but this text is also re-arranged in a new way. Accommodation has taken place if the article has been re-organized or new aspects have been integrated into existing information. A further example (by another participant) from the same study illustrates this accommodation process:

> *"Beyond biological causes, another potentially interesting reason is the interrelation of schizophrenia and the social environment. A study collected data about people's social class, their profession, education, income, and social situation (period of 2 years) as well as stress and personal living conditions that have negative psycho-social effects. The results were unambiguous: schizophrenic persons come from a low social class, have experienced inferior*

> *education, earn less money and are under considerable psycho-social strain. As a consequence, it may be concluded that a negative social environment contributes fundamentally to the onset of schizophrenia."*

People working on a wiki will not only improve their individual knowledge, but may also contribute to the development of collective knowledge at the same time. Thus, wikis support individual learning, and they may be an attractive tool for knowledge-building purposes, as described by Scardamalia and Bereiter. Collaboration with a wiki has the primary goal of advancing knowledge. Processes of externalization and internalization are going on all the time. In this way, both individual and collective knowledge are being developed constantly. The cognitive systems of individuals as well as the social system are continuously being enhanced and advanced, due to these processes of equilibration. All the people involved here can introduce information, and this information may be treated by each member of the community on an equal footing. A wiki community gets together when there is a need for jointly developing solutions to some problem which its members have in common. An idea that is introduced to the artifact will stimulate and inspire another community member to incorporate, develop, and improve this idea. In this way, concepts and ideas are developed, and the community and all the individuals involved can expand their knowledge.

## Conclusion

In this paper, we have elaborated on a framework in which knowledge building is defined as a co-evolution of cognitive and social systems. New knowledge may develop when people interact with each other via shared digital artifacts. It is assumed that collective knowledge becomes manifest in these shared artifacts. In order to illustrate these processes, we have demonstrated how they work, referring to two examples of social-software technologies: social-tagging systems and wikis. These allow different degrees of influencing and manipulating artifacts. Social tagging gives individuals only little scope for manipulating the whole artifact and its content. They can only add some new keywords or add a tag which has already been used. If many users have participated, resulting in a tag cloud which is already rather large, one individual contribution will probably not change this tag cloud very much. In such systems, it is not each individual's tagging behavior that will directly define the artifact. Instead, the artifact represents the accumulation of all the tags, giving each user equal impact. Regarding wikis, there are many opportunities for individual users to influence the artifact's content. They may, at their discretion, revise, change, or delete any part of a wiki article. If a user deletes some content, it will not be part of the shared artifact any more. The current state of a wiki does not represent the sum total of all these activities, as is the case with tagging systems. Instead, a modified wiki represents only the most recent interventions. This means that one person who modifies content in a wiki has the opportunity to make the whole text coherent. So in a wiki, a user can directly deal with incongruities.

To what extent a certain tool is conducive to knowledge building, depends very much on the potential of that technology to induce cognitive conflicts. In social tagging, productive conflicts are primarily caused by differences between a person's internal conceptual structure and that provided by the artifact. Individuals may solve this conflict mainly by changing their own individual knowledge structure. Wikis provide a high potential both for the development of cognitive conflicts and for solving them. The content of a wiki may differ fundamentally from its users' previous knowledge and understanding. If they want to improve the wiki text, they need to connect new content by re-organizing and re-conceptualizing it in order to adapt it to the content that already exists. This course of action will necessarily lead to the development of the shared artifact and to knowledge building.

To sum it up, this paper shows the potential of social software for learning and knowledge building. Social-software tools open up specific opportunities to combine individual and collective learning processes. Such tools are particularly relevant for knowledge-building purposes when they provide opportunities to influence and manipulate shared digital artifacts and when they have the potential to induce cognitive conflicts. Future developments of such technologies may benefit from taking these considerations into account.

## References


Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bereiter, C., & Scardamalia, M. (1996). Rethinking learning. In D. Olson & N. Torrance (Eds.), *Handbook of education and human development: New models of learning, teaching and schooling* (pp. 485-513). Cambridge, MA: Basil Blackwell.

Budiu, R., Pirolli, P., & Hong, L. (2009). Remembrance of things tagged: How tagging effort affects tag production and human memory. *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, pp. 615-624, New York, ACM Press.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Cress, U., & Kimmerle, J. (2007). A theoretical framework of collaborative knowledge building with wikis – a systemic and cognitive perspective. In C. A. Chinn, G. Erkens, & S. Puntambekar. (Eds.), *Proceedings* *of the 7th Computer Supported Collaborative Learning Conference* (pp. 153-161). New Brunswick, NJ: International Society of the Learning Sciences.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning, 3*, 105-122.

Désilets, A., Paquet, S., & Vinson, N. G. (2005). Are wikis usable? *WikiSym 2005 – Conference Proceedings of the 2005 International Symposium on Wikis*, pp. 3-15.

Fu, W. (2008). The microstructures of social tagging: A rational model. *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 229-238). New York: ACM.

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science, 32*, 198-208.

Hewitt, J., & Scardamalia, M. (1998). Design principles for distributed knowledge building processes. *Educational Psychology Review, 10*, 75-96.

Kimmerle, J., Moskaliuk, J., & Cress, U. (2009). Learning and knowledge building with social software. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Computer Supported Collaborative Learning Practices: CSCL2009 Conference Proceedings* (Vol. 1, pp. 459-468). International Society of the Learning Sciences (ISLS).

Köhler, A., & Fuchs-Kittowski, F. (2005). Integration of communities into process-oriented structures. *Journal of Universal Computer Science, 11*, 410-425.

Kolbitsch, J., & Maurer, H. (2006). The transformation of the web: How emerging communities shape the information we consume. *Journal of Universal Computer Science, 12*(2), 187-213.

Leuf, B., & Cunningham, W. (2001). *The wiki way. Quick collaboration on the web.* Boston: Addison-Wesley.

Luhmann, N. (1995). *Social systems.* Stanford: Stanford University Press.

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, Tagging Paper, Taxonomy Flickr, Academic Article, ToRead. *Proceedings of the 17th Conference on Hypertext and Hypermedia* (pp. 31-40). New York: ACM Press.

Moskaliuk, J., Kimmerle, J., & Cress, U. (2008). Learning and knowledge building with wikis: The impact of incongruity between people's knowledge and a wiki's information. *Proceedings of the International Conference for the Learning Sciences 2008, Vol. 2* (pp. 99-106). Utrecht, The Netherlands: International Society of the Learning Sciences.

Moskaliuk, J., & Kimmerle, J. (2009). Using wikis for organizational learning: Functional and psycho-social principles. *Development and Learning in Organizations, 23*(4), 21-24.

Nonaka, I. (1991). The knowledge-creating company. *Harvard Business Review, 69*, 96-104.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*, 14-37.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation.* Oxford: Oxford University Press.

Notari, M. (2006). How to use a Wiki in education: Wiki based effective constructive learning. *Proceedings of WikiSym'06 - 2006 International Symposium on Wikis* 2006, pp. 131-132.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures.* New York: Viking Press.

Polanyi, M. (1966). *The tacit dimension.* Garden City, NY: Doubleday.

Raitman, R., Augar, N., & Zhou, W. (2005). Employing wikis for online collaboration in the e-learning environment: Case study. *Proceedings of the 3rd International Conference on Information Technology and Applications, ICITA 2005* II, art. no. 1488944, pp. 142-146.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences 1*, 37-68.

Scardamalia, M., & Bereiter, C. (1996). Computer support for knowledge-building communities. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm.* Mahwah, NJ: Lawrence Erlbaum Associates.

Scardamalia, M., & Bereiter, C. (1999). Schools as knowledge building organizations. In D. Keating & C. Hertzman (Eds.), *Today's children, tomorrow's society: The developmental health and wealth of nations* (pp. 274-289). New York: Guilford.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In *Encyclopedia of education* (2nd ed., pp. 1370-1373). New York: Macmillan Reference.

Stahl, G. (2000). A model of collaborative knowledge-building. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences* (pp. 70-77). Mahwah, NJ: Erlbaum.

Wang, C.-M., & Turner, D. (2005). Extending the wiki paradigm for use in the classroom. *International Conference on Information Technology: Coding Computing, ITCC* 1, pp. 255-259.

Wasko, M. M., & Faraj, S. (2005).Why should I share? Examining knowledge contribution in networks of practice. *MIS Quarterly, 29,* 35-58.

# The Influence of Presentation Format and Subject Complexity on Learning from Illustrated Texts in Biology

Mareike Florax, Rolf Ploetzner, University of Education, Freiburg, Germany
mareike.florax@ph-freiburg.de, rolf.ploetzner@ph-freiburg.de

**Abstract:** In an experimental study, we investigated how learning from illustrated texts is influenced by presentation format (separated format, integrated format, and active integration) and subject complexity (low, medium and high complexity). Subject matter were biochemical processes of signal transmission in the human nervous system. A total of 180 students participated in the study. Results show that the active integration of texts and illustrations can improve the comprehension of highly complex material, whereas it can have inhibiting effects for less complex material. The results are discussed with respect to the process requirements associated with the different presentation formats.

## Introduction

Learning material often consists of different representations. A very common combination of representations is made up of expository texts and illustrations. The beneficial effects of combining texts and illustrations may be attributed to the fact that they complement each other with regard to their content (e.g., Ainsworth, Bibby & Wood, 2002), differ with regard to their computational efficiency (e.g., Larkin & Simon, 1987), or constrain each other's interpretations (Ainsworth, 2006). Although beneficial effects of learning from combinations of texts and illustrations have been frequently observed (e.g., Levie & Lentz, 1982), research has also identified certain difficulties that learners may encounter when processing such combinations. While some of these difficulties seem to be related to the design of the learning material, others seem to be intrinsic to the learners such as their prior knowledge and learning preferences. For instance, it is assumed that the design of learning material may pose unnecessary high processing demands on the learners which, in turn, can lead to an overburdening of their cognitive capacities (e.g., Sweller, van Merriënboer & Paas, 1998). With regard to difficulties intrinsic to the learners, research indicates that particularly those learners with only little prior knowledge may experience difficulties in relating different representations to each other. This seems to be especially true if the learning material is of high complexity (e.g., Bodemer & Faust, 2006; Kozma, 2003). Research also indicates that many learners engage in rather shallow information processing when learning from texts and illustrations (e.g., Peek, 1993).

In all the cases mentioned, learners do not take advantage of the synergetic effects that texts and illustrations may provide them and, as a consequence, fail to construct a coherent mental model of the information presented (cf. Ainsworth et al., 2002). In order to be able to fully exploit the potential of combined texts and illustrations, learners seem to need instructional support (cf. Ainsworth et al., 2002; Bodemer et al., 2004; Hegarty, Carpenter & Just, 1991). At least two approaches aiming at such support can be distinguished. One approach, the *design-oriented approach*, focuses on improving the design of learning material. A second approach, the *engagement-oriented approach*, aims at inducing appropriate mental processes in the learners. In three studies, Bodemer et al. (2004, 2005) demonstrated that learning can be improved by encouraging learners to spatially integrate texts and illustrations by themselves. Because these studies revealed that higher learning gains were associated with higher subject complexity, Bodemer et al. (2004, 2005) assumed that the higher the subject complexity, the more advantageous is learning by integrating texts and illustrations. However, since each study investigated learning in a different domain, the results of the three studies were difficult to compare.

In this paper, we present an experimental study in which we varied the presentation format as well as the subject complexity within the same domain of biology. We begin by outlining the theoretical assumptions and empirical findings within the design-oriented and engagement-oriented approaches. The role of complexity in learning from combined texts and illustrations is then explored. Thereafter, the experimental study and its results are described and discussed.

### The Design-Oriented Approach

Research on how to design combinations of texts and illustrations has been highly influenced by two theories: Richard Mayer's *Cognitive Theory of Multimedia Learning* (CTML; Mayer, 2001, 2005) and John Sweller's *Cognitive Load Theory* (CLT; Ayres & Sweller, 2005). CTML takes a process-oriented perspective on learning from combined texts and illustrations. Based on Paivio´s (1986) dual coding theory, and the assumption that human working memory is limited in its capacity (Baddeley, 2003), CTML postulates the following cognitive processes required for successful learning from multimedia: a) selecting relevant information in each presentation, b) organizing the selected information in either a verbal or a pictorial mental model, and c) integrating the verbal and pictorial models as well as prior knowledge by means of mapping processes between corresponding information (Mayer, 2005). Like other conceptualizations of learning from multimedia, CTML considers the integration of verbal and pictorial information to be pivotal for constructing a mental model, which in turn is considered to be important for deep understanding. On the basis of his theory, Mayer (2001) advises that texts and illustrations should be presented in spatial proximity rather than spatially separated. Essentially, it is assumed that spatial proximity facilitates the mental integration of textual and pictorial information. In numerous studies, Mayer and colleagues (e.g., Mayer, 1989; Moreno & Mayer, 1999) were able to demonstrate that a spatially contiguous presentation format leads to more successful learning than does a spatially separated presentation format. This empirical finding has been termed the *spatial contiguity effect*.

As in Mayer's (2001) CTML, it is also assumed in the resource-oriented CLT that human working memory is of limited capacity (e.g., Ayres & Sweller, 2005). According to CLT, three different types of cognitive load may burden working memory during learning: (1) extraneous cognitive load, which results from unnecessary cognitive processes induced by the inappropriate design of learning material; (2) intrinsic cognitive load, which is imposed by subject complexity which, in turn, is determined by the interactivity of elements in the learning material as well as the learners´ prior knowledge; and (3) germane cognitive load, which is related to learning processes such as the construction of concepts and schemata (cf. Ayres & Sweller, 2005).

For a long time, research on CLT has focused on reducing extraneous cognitive load by improving the design of learning material (van Merriënboer & Sweller, 2005). For instance, it is assumed that texts and illustrations which are spatially split induce unnecessary processes such as visual search processes. These processes contribute to extraneous cognitive load and may therefore impede learning. In contrast to Mayer's (2001) CTML, CLT recommends not only to present texts and illustrations together in spatial proximity, but to spatially integrate segments of the texts into the illustrations, resulting in a so-called integrated format. Envision, a text directly positioned above an illustration. From the perspective of CTML, this would correspond to a spatially proximate presentation format because both the text and illustration are presented close to each other. From the perspective of Sweller's CLT, however, this would still correspond to a spatially separated presentation format because the text is not integrated into the illustration. Sweller and colleagues have demonstrated that an integrated format leads to more successful learning than a spatially separated format (e.g., Cerpa, Chandler & Sweller, 1996). They termed this empirical finding the *split-attention effect*. Research also revealed that an integrated format is especially beneficial when the subject complexity is high (e.g., Pollock, Chandler & Sweller, 2002; Ginns, 2006).

A further design principle aimed to reduce extraneous cognitive load and to support the mental integration of verbal and pictorial information is to make the corresponding information in each presentation explicit. This can be achieved by visual indicators such as colour-codes (e.g., Bodemer et al., 2004; Tabbers, Martens & van Merriënboer, 2004), highlighting (e.g., Jamet, Gavota & Quaireau, 2008), or dynamic linking. For example, van der Meij and de Jong (2006) found that learning from dynamically linked presentations is especially beneficial when the subject complexity is high.

Whether corresponding information is made explicit by spatial proximity or by visual indicators, in both cases it is completely up to the learners as to whether or not they engage in the desired selection, organization, and integration processes. Although well-designed learning material may free the learners from unnecessary cognitive processes, thereby making more cognitive capacity available for learning, it is unknown to which degree learners actually take advantage of this capacity. Perhaps learning could be further improved by systematically encouraging the learners to engage in the desired learning processes. Furthermore, various studies indicate that providing learners with well-designed learning material may even act to inhibit learning, since some learners undervalue the complexity and difficulty of the learning material, and therefore exhibit only shallow information processing (e.g., McNamara, Kintsch, Songer & Kintsch, 1996; Schnotz & Rasch, 2005).

## The Engagement-Oriented Approach

In contrast to the design-oriented approach, the engagement-oriented approach encourages learners to more actively process the information presented. Various theories of learning, especially constructivist theories, emphasize that the learners' engagement is crucial to successful learning. For example, in Wittrock´s (1992) theory of generative learning, it is assumed that learners do not passively take up and store information, rather they actively search for and process information in order to construct knowledge. Learning thus takes place by actively establishing relations between different pieces of information that are presented to the learners, as well as by establishing relations between new information and prior knowledge. Accordingly, Wittrock's (1992) theory implies that understanding depends on the mental processes actively carried out by the learners.

Various methods have been developed in order to induce and support the learners' engagement in active information processing. Two examples which have been proven to facilitate learning are the production of self-explanations (e.g., Chi et al., 1989; Renkl, 2002) and the visualisation of information present in texts (e.g., van Meter, Aleksic, Schwartz & Garner, 2006). Peek (1993) also recommends instructional interventions which lead to an external and observable product. For instance, various instructions aim at inducing the relevant learning processes by asking the learners to organize textual and pictorial components on the computer screen (e.g., Bodemer et al., 2004).

Bodemer et al. (2004, 2005) developed an instructional method which takes all mentioned aspects into account. They termed this method *active integration*. Initially, learners were presented spatially separated texts and illustrations on a computer screen. They were then asked to move segments of the text to their referential areas in the illustration using the drag and drop method. After completing the task, the learners had constructed a spatially integrated presentation format by themselves. In three experimental studies, Bodemer and colleagues were able to demonstrate that learners who integrated texts and illustrations themselves outperformed learners who were provided with a spatially integrated or a spatially separated presentation format (Bodemer et al., 2004; 2005; Bodemer & Faust, 2006).

Because higher learning gains were attained with higher subject complexity in the studies of Bodemer et al. (2004, 2005), the authors assumed that learning by integrating texts and illustrations may be more successful as the subject complexity increases. This assumption is in line with findings concerning the split-attention effect (cf. Ginns, 2006): while learning from an integrated format is most beneficial if the subject complexity is high, it leads to only small improvements if the subject complexity is low. However, because learning encompassed a different subject domain (physics, statistics and mechanics) in each study of Bodemer et al. (2004, 2005), the results of their studies were difficult to compare.

In the following, we present an experimental study which investigates how presentation format and subject complexity influence learning from combined texts and illustrations. In this study, subject complexity is varied systematically within the same subject domain. According to Sweller and Chandler (1994), subject complexity denotes the number of interacting elements that have to be processed simultaneously by the learners. Increases in complexity were achieved by successively adding interacting elements to the learning material.

## Experiment

### Design

Two factors, each with three variations, were considered: (1) *presentation format* of combined texts and illustrations (separated format, integrated format, active integration) and (2) *subject complexity* (low, medium, high). Dependent variables were retention and comprehension of the learning material. The learners´ working memory capacity was assessed and taken into account as a covariate. Furthermore, the learners´ spatial abilities were also assessed and taken into account as a covariate. Because learning time was not fixed, it was recorded as a further covariate.

### Hypotheses

In accord with Bodemer et al. (2004, 2005), we expected that active integration facilitates learning from combined texts and illustrations more than learning from an integrated format. This hypothesis relies on the assumption that active integration induces and supports the learners' engagement. Because such engagement processes are more important with respect to comprehension than to retention, more substantial effects were expected concerning comprehension. In accord with Ayres and Sweller (2005), as well as Mayer (2001), we expected learning from an integrated format to be more beneficial than learning from a separated format; i.e., we expected the split-attention effect. With regard to subject complexity, we expected active integration to be more beneficial, the higher the subject complexity. More substantial effects of subject complexity were expected with respect to comprehension than with respect to retention. These hypotheses rely on the assumption that during learning from complex material learners need more encouragement and support to systematically and comprehensively process the material.

### Material

The subject domain was information processes of the human nervous system. Various biochemical processes related to the transmission of signals between neurons were verbally described and graphically illustrated in such a way that the texts and illustrations complemented each other. For each group, the computer-based learning material consisted of an explanatory text and an accompanying illustration which were presented together on a single computer screen. In order to systematically increase the subject complexity, interacting textual and pictorial information was successively added to the material. Learning material of low complexity addressed the development of resting potentials within the human nervous system (cf. Figure 1). Learning material of medium complexity additionally described the development of action potentials. Processes involved in resting potentials and those involved in action potentials influence each other. Therefore, the processes described in the material of medium complexity did not only complement the processes described in the material of low complexity but the former interacted with the latter. Learning material of high complexity additionally described processes of inhibiting action potentials (cf. Figure 2) which, again, interacted with processes described in the material of low and medium complexity.

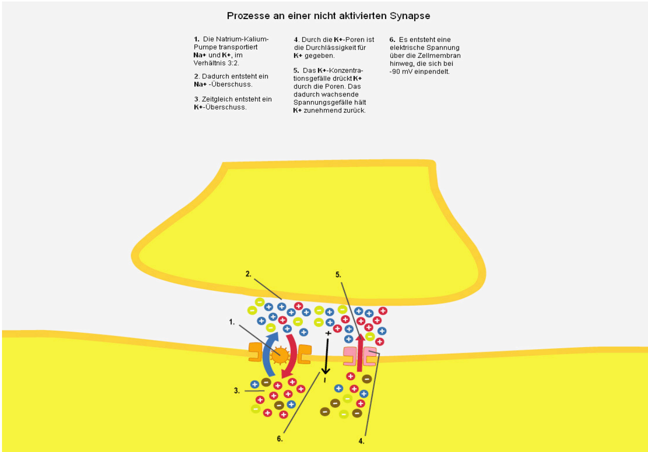



Figure 1. Learning material of low complexity in a separated format.

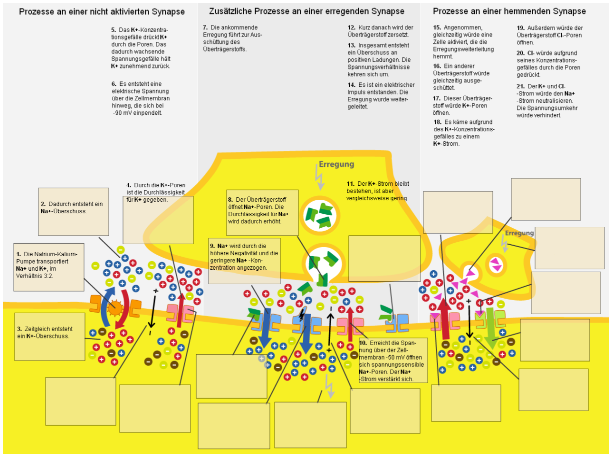



Figure 2. Learning material of high complexity with active integration.

Figure 1 also shows an example of the separated format (SF). Each text segment and corresponding part of the illustration were numbered according to the reading sequence. In the integrated format (IF), the text segments were placed next to the corresponding parts of the illustration. Active integration (AI) began with a separated format (cf. Figure 2); participants were then required to drag and drop the given text segments onto the appropriate part of the illustration. After having arranged all text segments, participants received feedback. If segments were placed incorrectly, the participants were given two opportunities to revise their integration. In the end, the participants were presented with the correct integrated format.

All participants received a short introductory text about the general structures and processes within the human nervous system. On the basis of four multiple-choice questions, participants who did not properly process the introductory text were excluded from the study; at least two of the the four questions had to be answered correctly. To familiarize participants in the active integration group with dragging and dropping text segments, they received an introduction to this functionality.

A pre-test of the participants´ prior knowledge was made up of eight multiple-choice questions which addressed the learning material. Three overlapping post-tests were constructed addressing respectively the low, medium and high complex learning material. All post-tests were comprised of questions relating to retention and comprehension. Both types of questions were constructed in such a way that the correct answers required the mental combination of verbally and pictorially presented information. Whereas questions on retention asked for the recall of information explicitly presented in the learning material, questions on comprehension demanded the participants to draw inferences from the information presented. The post-test for low complex material consisted of eight questions on retention and five questions on comprehension. The questions on retention were identical to the questions used in the pre-test. The post-test for medium complex material was comprised of eleven questions on retention and seven questions on comprehension. In the case of retention, five questions were identical to questions of the post-test for low complex material and six questions were new. In the case of comprehension, three questions were identical to questions on the post-test for low complex material and four questions were new. The post-test for high complex material was also comprised of eleven questions on retention and seven questions on comprehension. In the case of retention, seven questions were identical to questions found on the post-test for medium complex material and four questions were new. In the case of comprehension, five questions were identical to questions on the post-test for medium complex material and two questions were new. Pre- and post-tests were administered on the computer. Each item answered correctly was scored with one point.

With respect to the covariates, working memory capacity was assessed by means of the subtest "memorizing digits" from the Wilde Intelligence Test (Jäger & Althoff, 1983). Spatial abilities were assessed by means of three subtests of the System for the Assessment of Performance (Horn, 1983).

## Procedure

Participants were investigated in groups of no more than three; they worked, however, individually on the different material. Sessions were structured into five phases. During the first phase, participants took the memory capacity test and the spatial ability test. During the second phase, participants read the introductory text and answered the pertaining questions. They were allowed to keep the text until the post-test started. During the third phase, learners took the pre-test. The learning material was processed in the fourth phase. The amount of learning time could be chosen individually and was recorded by the computer. Finally, during the fifth phase, participants worked on the post-test.

## Participants

Overall, 180 students from two different universities in Freiburg, Germany, were randomly assigned to the nine experimental groups. Students of medicine, psychology and the natural sciences did not participate in the study. Two participants were excluded from the study due to insufficient processing of the introductory text. Five participants were excluded due to high prior knowledge: they answered more than 25% of the pre-test questions correctly. Four participants were excluded due to technical problems. Overall, 169 participants remained (age: $M = 22.3$, $SD = 2.4$; gender: 58 males, 111 females).

## Results

### Prior Knowledge

On average, the participants' prior knowledge was low ($M = .34$ (4.3%), $SD = .62$) and there were no significant differences between groups with respect to prior knowledge ($F(8,160) = 1.18$, *n.s.*). There was also no significant correlation between performance on the pre- and post-test ($r = .03$, *n.s.*). Therefore, prior knowledge was not further considered in the analysis.

### Covariates

An analysis of variance with the factors presentation format and subject complexity on learning time revealed a statistically significant effect for subject complexity ($F(2,160) = 130.30$, $p < .01$, $\eta^2_{part} = .62$). There was no statistically significant effect for presentation format ($F(2,160) = 1.13$, *n.s.*). The interaction between the two factors was also not statistically significant ($F(4,160) = 1.55$, *n.s.*). Post-hoc tests indicated that learning from low complex material took significantly less time than learning from medium complex material (Games-Howell, $MD = -5.01$, $SE = .44$, $p < .001$), which, in turn, took significantly less time than learning from highly complex material (Games-Howell, $MD = -3.87$, $SE = .63$, $p < .001$). On average, participants exhibited high spatial abilities ($M = 6.8$ (75.5%), $SD = 1.1$) and medium working memory capacity ($M = 100.9$ (52.3%), $SD = 10.1$). Analyses of variance showed no significant differences between groups, neither for spatial abilities ($F(8,160) = 1.27$, *n.s.*) nor for working memory capacity ($F(8,160) = 1.01$, *n.s.*).

### Learning Performance

With respect to highly complex material, descriptive data shows that active integration led to more successful learning than learning from an integrated format, which in turn was superior to learning from a separated format (cf. Table 1). The differences between these conditions were larger for comprehension than for retention. While the descriptive trends for medium complex material were similar to those of highly complex material, the descriptive trends for low complex material reversed.

A multivariate analysis of covariance (MANCOVA) on the dependent variables retention and comprehension was computed; it included the factors presentation format and subject complexity, as well as the covariates spatial ability, working memory capacity and learning time. Overall, it revealed no significant influence of learning time ($F(2,156) = 2.26$, *n.s.*) on post-test results, but significant influences of spatial ability ($F(2,156) = 3.24$, $p < .05$) and working memory capacity ($F(2,156) = 5.90$, $p < .01$). Furthermore, a statistically significant effect was found for subject complexity ($F(4,312) = 3.78$, $p < .01$). The analysis revealed no statistically significant effect for presentation format and no statistically significant interaction effect between subject complexity and presentation format. The main multivariate effect of subject complexity can be attributed to statistically significant differences with respect to retention ($F(2,157) = 3.43$, $p < .05$).

Table 1: The means and standard deviations of the relative solution frequencies in the post test.

| | | Low complexity | Medium complexity | High complexity |
|---|---|---|---|---|
| **Separated format** | Retention | 63.1% (23.5%) | 72.7% (19.9%) | 54.5% (29.1%) |
| | Comprehension | 40.0% (24.3%) | 45.1% (28.3%) | 42.0% (24.5%) |
| | Overall | 54.2% (20.5%) | 62.0% (20.7%) | 49.7% (25.1%) |
| **Integrated format** | Retention | 62.5% (27.0%) | 66.0% (23.2%) | 59.1% (23.3%) |
| | Comprehension | 38.9% (29.4%) | 45.1% (20.4%) | 51.4% (22.0%) |
| | Overall | 53.4% (25.3%) | 57.9% (20.1%) | 56.1% (20.3%) |
| **Active integration** | Retention | 56.3% (31.0%) | 60.5% (27.7%) | 61.5% (28.3%) |
| | Comprehension | 28.9% (24.0%) | 44.3% (21.7%) | 55.5% (23.1%) |
| | Overall | 45.7% (25.8%) | 54.2% (22.5%) | 59.2% (24.1%) |
| **Overall** | Retention | 60.7% (26.9%) | 66.3% (24.0%) | 58.7% (26.5%) |
| | Comprehension | 36.1% (26.0%) | 44.8% (23.4%) | 49.7% (23.3%) |
| | Overall | 51.3% (23.8%) | 58.0% (21.0%) | 55.0% (23.0%) |

In order to further explore the data, we conducted a trimmed means analysis according to Wilcox (1998). Based on the results in the post-test, the lower 10% as well as the upper 10% of the participants were removed from each group. In the analysis of trimmed means, the covariates learning time, spatial ability, and working memory capacity showed less influence on post-test results than in the analysis of untrimmed means. The most important difference between the two analyses was that the interaction between presentation format and subject complexity reached statistical significance with respect to comprehension ($F(4,121) = 2.52$, $p < .05$, cf. Figure 3). According to a post-hoc simple effects analysis, this interaction results from the changing effects of active integration ($F(2,123) = 7.34$, $p < .001$): The active integration of learning material inhibits learning when subject complexity is low, but facilitates learning when subject complexity is high. The simple effects analysis did not reach statistical significance with respect to the seperated and integrated presentation formats.

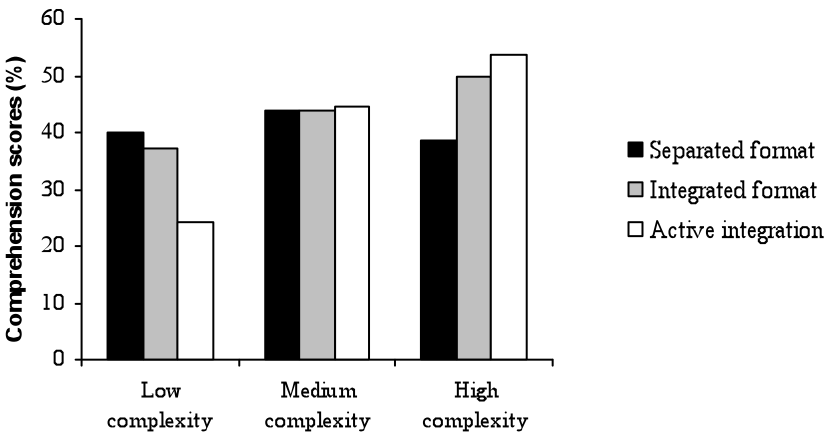



Figure 3. Comprehension scores for the three formats at different levels of complexity.

## Discussion

Whether the active integration of texts and illustrations as proposed by Bodemer et al. (2004, 2005) acts to inhibit or to facilitate learning is dependent upon the subject complexity. However, it was only after reducing the variances within groups, that the expected interaction between presentation format and subject complexity was statistically confirmed with respect to comprehension. When subject complexity was high, actively integrating texts and illustrations was of greater benefit than learning from a separated or an integrated format. The active integration of text and illustrations inhibited learning when subject complexity was low. What could have caused this inhibition? It could be that the participants underestimated the difficulty of learning from the low complex material. In the low complex material, the amount of textual and pictorial information was rather small and both types of information were clearly arranged. Requesting the participants to integrate this information might not have led to the intended effortful engagement leading to a deeper understanding of the material, but rather to a more or less “mindless” execution of the drag-and-drop-task (cf. Salomon, 1983). Overall, our findings support, or to say it more carefully, do not contradict the hypothesis of Bodemer et al. (2004, 2005) that learning by integrating texts and illustrations is most successful when the subject complexity is high.

Like Martin-Michiellot and Mendelsohn (2000), as well as Bodemer et al. (2004), we were not able to replicate the split-attention effect - not even with respect to highly complex material. This finding is in contradiction to many studies, especially those by Sweller and colleagues (e.g., Cerpa, Chandler & Sweller, 1996). A comparison of the learning material used, however, reveals two differences. First, in the spatially separated learning material used in many other studies, a more or less unstructured text was combined with an illustration. In many cases, the relevant areas within the illustration were not clearly distinguished. Thus, it was up to the learners to identify meaningful assertions in the text and relevant parts of the illustration, and to then appropriately relate them to each other. In these studies, the separated and integrated formats differed not only with regard to spatial proximity, but with regard to other factors as well such as text segmentation and picture labelling (e.g., Chandler & Sweller, 1991). In contrast, the text in our learning material was always structured into meaningful segments and the relevant areas within the illustration were indicated. It is therefore plausible that our material imposed smaller extraneous cognitive loads on the learners than the material used in other studies. As a consequence, the chance that a split-attention effect would occur was also smaller. From a theoretical point of view, the finding that an integrated format does not lead to better learning than a well-structured separated format supports an explanation put forward by Erhel and Jamet (2006): the beneficial effects of integrated formats result from the explication of relations between texts and illustrations rather than from spatial proximity. A second difference is the design of the pictorial material employed. While Sweller and colleagues (e.g., Sweller & Chandler, 1994) often made use of simple black-and-white line drawings, we employed coloured schematic illustrations comprised of many visual elements. Both colours, as well as the number of visual elements, increase the visual richness of an illustration (cf. Peek, 1993). This difference also could have contributed to a convergence of the extraneous cognitive load imposed on the learners by the separated and integrated formats.

Overall, the differences addressed above indicate that the roles of segmenting text, pointing out relevant parts of illustrations, and designing visual material are underspecified in the context of the split-attention effect. It appears that the distinction between a separated format and an integrated format might not correspond to a clear-cut dichotomy. Martin-Michiellot and Mendelsohn (2000) assume that texts and illustrations can be designed and arranged in such a way that they do not yield an integrated format, but nevertheless support the learner in relating and mentally integrating both presentations. In a recent study, Florax and Ploetzner (in press) found evidence for this assumption: learning from segmented texts and spatially separated pictures improved learning as much as learning from spatially integrated texts and pictures.

The reported study is a first step to clarify the role of active integration for learning from material of different complexity. According to our observations, active integration should only be employed when the subject complexity is high, otherwise it might hinder learning. For example, in the realm of academia, where learning material is usually of high complexity, active integration has the potential to improve learning.

## References


Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 138-198.

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences, 11*, 25-61.

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135-146). New York: Cambridge University Press.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839.

Bodemer, D., & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior, 22*, 27-42.

Bodemer, D., Ploetzner, R., Bruchmüller, K., & Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instructional Science, 33*, 73-95.

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction, 14*, 325-341.

Cerpa, N., Chander, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research, 15*, 345-367.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Erhel, S. & Jamet, E. (2006). Using pop-up windows to improve multimedia learning. *Journal of Computer Assisted Learning, 22*, 137-147.

Florax, M., & Ploetzner, R. (in press). What contributes to the split-attention effect? The role of text-segmentation, picture labelling, and spatial proximity. *Learning and Instruction*.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*, 511-525.
Hegarty, M., Carpenter, P. A., & Just, M. A. (1991). Diagrams in the comprehension of scientific texts. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 641-668). New York: Longman.
Horn, W. (1983). *Leistungsprüfsystem [System for the assessment of performance]* (2nd ed.). Göttingen: Hogrefe.
Jamet, E. & Gavona, M., & Quaireau, C. (2008). Attention guiding in multimedia learning, *Learning and Instruction*, 18, 135-145.
Jäger, A. O., & Althoff, K. (1983). *Der WILDE-Intelligenztest (WIT) [The WILDE intelligence test]*. Göttingen: Hogrefe.
Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction, 13*, 205-226.
Larkin, J. H., & Simon, H. H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.
Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal, 30*, 195-232.
Martin-Michiellot, S., & Mendelsohn, P. (2000). Cognitive load while learning with a graphical computer interface. *Journal of Computer Assisted Learning, 16*, 284-293.
Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81*, 240-246.
Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York: Cambridge University Press.
McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.
Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.
Paivio, A. (1986). *Mental representations - A dual coding approach*. Oxford: Oxford University Press.
Peeck, J. (1993). Increasing picture effects in learning from illustrated text. *Learning and Instruction, 3*, 227-238.
Pollock, E., Chandler P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*, 61-86.
Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanation. *Learning and Instruction, 12*, 529-556.
Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist, 18*, 42-50.
Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development, 53*, 47-58.
Sweller, J., & Chander, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185-233.
Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.
Tabbers, H., Martens, R., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*, 71-81.
Van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*, 199-212.
Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147-177.
Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology, 31*, 142-166.
Wilcox, R. M. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53*, 300-314.
Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist, 27*, 531-541.

## Acknowledgments

The research reported in this paper was supported by the Deutsche Forschungsgemeinschaft (DFG) under contract PL 224/11-1.

# An invisible preference for intrinsic motivation in Computer-Mediated Communication

Bart Rienties, Bas Giesbers, Dirk Tempelaar, Mien Segers, Wim Gijselaers, Maastricht University, Tongersestraat 53, 6200 MD Maastricht, The Netherlands.
Bart.rienties@maastrichtuniversity.nl, s.giesbers@maastrichtuniversity.nl, d.tempelaar@maastrichtuniversity.nl, m.segers@maastrichtuniversity.nl, w.gijselaers@maastrichtuniversity.nl.

**Abstract:** A large number of studies in CMC have assessed how social interaction, learning processes and outcomes are intertwined. Although recent research findings indicate that learners differ with respect to the amount and type of discourse contributed in virtual settings, little is known about the underlying causes and its consequences explaining differences between participants' contributions to discourse. The present research investigates how motivational orientation of a learner influences the interaction patterns with other learners in a virtual network.

Our research among 100 participants in six virtual teams indicates that three sub-groups were formed within each virtual network. These subgroups were generated by a K-means cluster analysis of academic motivation measured by AMS. Extrinsically motivated learners prefer to connect to intrinsically motivated learners. Intrinsically motivated learners prefer to discuss mainly among themselves, implying that extrinsically motivated learners will receive less feedback and discourse possibilities from other members within the virtual network.

## Introduction

In recent years, the attention for virtual collaborative learning seems to be fuelled by two separate, yet mutually enforcing developments: First, the availability of increasing possibilities of Information Communication Technologies (ICT) provide enhanced support for collaboration (Bromme, Hesse, & Spada, 2005; Resta & Laferrière, 2007; Schellens & Valcke, 2005). Second, growing amounts of evidence have become available showing that collaboration can enrich student learning through interaction (Jonassen & Kwon, 2001; Lindblom-Ylänne, Pihlajamäki, & Kotkas, 2003; Van den Bossche, Gijselaers, Segers, & Kirschner, 2006). In general, it can be said that virtual collaborative learning is built on the assumption that ICT has the power to provide a rich learning experience by using a variety of learning methods (Beers, Boshuizen, Kirschner, & Gijselaers, 2007; Giesbers, Rienties, Gijselaers, Segers, & Tempelaar, 2009; Jonassen & Kwon, 2001; Resta & Laferrière, 2007).

Despite the learning possibilities created by ICT-tools, recent findings in research on computer-mediated communication (CMC) indicate that interaction and contributions made to interaction depend on a variety of factors. Not every learner contributes equally to others. It has been found that learners who are similar with respect to educational background and prior knowledge nevertheless contribute differently to discourse (Caspi, Chajut, Saporta, & Beyth-Marom, 2006; De Laat, Lally, Lipponen, & Simons, 2007; Martens, Gulikers, & Bastiaens, 2004; Rienties, Tempelaar, Van den Bossche, Gijselaers, & Segers, 2009). An overall finding in research is that the majority of interactions and contributions within online courses can be attributed to a small number of learners. While this research has demonstrated the existence of this phenomenon, the obvious question is how these differential patterns between learners can be explained.

Several studies have examined this phenomenon by using social network analysis to explain interaction patterns in CMC. Social network analysis (SNA) provides powerful tools to analyze how people interact over a given period of time (Hurme, Palonen, & Järvelä, 2007; Martinez, Dimitriadis, Rubia, Gomez, & de la Fuente, 2003; Rienties et al., 2009). It considers whether certain individuals are central in networks or at the peripheries, and how interactions between individuals may change over time. Research findings have indeed revealed that some learners are more central in the social network than other learners (Hurme et al., 2007; Russo & Koesten, 2005). It has been found that learners who were central in a social network received and also contributed more messages than other learners. In addition, Russo and Koesten (2005) found that learners who were central in the social network had better cognitive learning outcomes.

Although recently several researchers have identified that some participants are more likely than others to be in the centre of networks, they could not explain the underlying mechanisms of these social interaction patterns. The present study aims to fill in this gap, by examining the underlying causes that explain differential contributions to social networks. We address the question why some learners receive relatively more replies to their contributions in discourse while others do contribute but get only limited response from others. So the issue is to what extent is it a coincidence that some learners become central contributors? In this article, we will investigate what the "invisible" mechanisms in social interaction are that result in learners of virtual networks being central or learners being on the outer fringe of a social network.

## An invisible hand in social interaction in CMC: motivation

In most virtual networks learners are geographically separated and dispersed over many settings. In collaborative settings, learners have to construct meaning and co-construct knowledge in a virtual setting. However, participation and making contributions to discourse in collaborative settings cannot be taken for granted (Bromme et al., 2005; Kirschner, Beers, Boshuizen, & Gijselaers, 2008). In particular when learners are interacting using lean ICT-tools like discussion forums or WIKIs, establishing a critical mass of interaction whereby participants contribute actively to cognitive discourse is troublesome (Caspi, Gorsky, & Chajut, 2003; Schellens & Valcke, 2005). Some learners are more inclined to start and actively contribute to a discussion than others. Other learners might prefer to wait for a while before contributing to a discussion, in particular when the members of the virtual network are seeking for effective working and learning strategies (Beers et al., 2007; Kirschner et al., 2008).

Within CMC, several researchers have tried to influence the interaction patterns among learners by (re-)scaffolding the learning process by designing scripts (Weinberger & Fischer, 2006), adjusting the degree of social presence of ICT tools (Giesbers et al., 2009; Jonassen & Kwon, 2001; Tu & McIsaac, 2002) or regulating the interaction processes (Kirschner et al., 2008). For example, by increasing the regulation of interaction processes, Beers et al. (2005) found that interaction among participants could be enhanced. By establishing argumentative scripts, learners contributed more argumentative discourse than when other scripts were used (Weinberger & Fischer, 2006). Nonetheless, individual differences to contributions to discourse still persist when redesigning the learning environment. Limited research has been conducted how differences in individual traits influence the interaction patterns of learners in networks.

One of the explanations for these individual differences lies in the motivation of learners to contribute to the virtual network. Recent research highlights indeed that motivation has a strong influence on how learners contribute to discourse in online settings (Järvelä, Järvenoja, & Veermans, 2008; Martens et al., 2004; Rienties et al., 2009; Veermans & Lallimo, 2007; Yang, Tsai, Kim, Cho, & Laffey, 2006). For example, Yang et al. (2006) conducted a survey among 250 respondents of eleven online educational psychology courses and found that goal-oriented motivation positively influences social presence among peers, that is the perception that emotions can be shared using CMC. Veermans and Lallimo (2007) found that messages contributed by motivated students demonstrate a richer variety of topics. Järvelä et al. (2008) found that students in the face-to-face setting reported more (favourable) learning goals and less performance goals relative to students in virtual settings.

The present research builds further on these findings by examining how motivation affects the creation, development and evolution of links between learners in virtual networks. As motivation is a multidimensional and multilevel construct (Boekaerts & Minnaert, 2006), a wide variety of definitions and instruments are discussed and used in educational psychology research. We adopt the concept of motivation developed by Deci and Ryan (1985), where to be motivated means to be moved to do something. The degree of self-determination of learners might explain why some learners contribute more to discourse in CMC than others. As a consequence, it is expected that some learners contribute more to discourse in CMC than others, given their motivation. However, focusing only on the level of motivation ignores the underlying attitudes and goals the learner has in order to pursue an action or goal (Deci & Ryan, 1985).

*Evolution of Social Networks*

The present study considers the way learners interact in virtual networks as social network interactions. According to Newman (2003), "[a] social network is a set of people or groups of people with some pattern of contacts or interactions between them". Within educational psychology, limited research has been conducted to understand dynamic social network interactions. According to network theorists, there are two important conditions that determine how social networks evolve: 1) the stability of the number of nodes (i.e. participants in virtual network); 2) the (in)equality of characteristics of nodes in the network (Barabási & Albert, 1999; Erdős & Rényi, 1960; Newman, 2003). In case the number of participants in a social network grows continuously (e.g. Wikipedia, Facebook), being among the first participants in the social network might imply that one is more likely to be connected to others than when one has recently joined a social network. In contrast, in an online course (as in most classes), the number of learners is mostly pre-determined and relatively stable. Therefore, a straightforward assumption from network theory would be that the social network of an online course will develop and evolve according to random graph theory (Barabási, 2002; Erdős & Rényi, 1960). In random graph theory, learners connect to other learners in a network with a more or less equal probability.
*H1:Learners in a virtual network will have an equal amount of connections to all other learners.*

If hypothesis 1 has to be rejected in our setting, then learners in virtual networks do not connect to other learners in line with the random graph theory. A crucial assumption of random graph theory is that people in the social network are perceived by others as equal (Erdős & Rényi, 1960; Newman, 2003). However, in line with the second condition when nodes (i.e. learners) have a specific preference to connect to some type of nodes, the network will not develop according to random graph theory (Barabási & Albert, 1999; Newman, 2003). Several researchers have indicated that learners in online settings differ with respect to prior knowledge, expertise and motivation when they become member of a virtual network (Järvelä et al., 2008; Martens et al., 2004; Rienties et al., 2009; Yang et al., 2006). When learners in a virtual network become aware that interacting with some learners who have a trait (e.g. intrinsic motivation, large knowledge base, expertise) that is (perceived to be) beneficial, these learners might be more interesting to interact with (Martens et al., 2004).

As intrinsically motivated learners are more inclined to contribute to discourse than extrinsically motivated learners, in particular with regard to higher cognitive discourse (Rienties et al., 2009), they possess crucial characteristics for distance learning. Superior contributions to discourse at a higher cognitive level might bring them a positive (expert) reputation in the virtual network. Other learners might be more willing to contribute to a learner who is perceived to be motivated and has some expert knowledge. In addition, as extrinsically motivated learners will perceive a lack of external regulation in distance learning, they might direct their attention more towards intrinsically motivated learners. In other words, intrinsically motivated learners lead the discourse development within the virtual network, thereby providing the desired external regulation to extrinsically motivated learners. This will imply that most learners will be connected to intrinsically motivated learners, as phrased in our second and third research hypotheses: *H2: Extrinsically motivated learners are more likely to interact with intrinsically motivated learners than with other extrinsically motivated learners; H3: Intrinsically motivated learners are more likely to interact with other intrinsically motivated learners than with extrinsically motivated learners.*

## Method

### Setting

The present study took place in an online summer course for prospective bachelor students of an International Business degree program at an Institute for Higher Education in the Netherlands (Rienties, Tempelaar, Waterval, Rehm, & Gijselaers, 2006). This online course was given over a period of six weeks in which learners were assumed to work for 10-15 hours per week. The participants never met face-to-face before or during the course and had to learn using the virtual learning environment "on-the-fly". In our setting, learners participated in virtual networks within a collaborative learning environment using discussion forums and announcement boards. During six weeks, learners had to collaborate together on solving six tasks through a problem-based learning method. No obligatory meetings were scheduled. The results of three interim-tests and a final summative test combined with graded participation in the discussion forums were used to make a pass-fail decision. Learners who passed the course received a certificate.

### Participants

In total 100 non-Dutch participants were randomly assigned in six networks. Data were analysed for those individuals who actually posted at least once a reaction in the discussion forum. We found that 18 learners, although registered for this course, never posted a contribution to the discussion forums. The 82 participants who posted at least once a reaction in the discussion forum were selected for our analysis. The six networks had an average of 13.66 members (SD = 2.16, range = 11-17) per network. The average age was 19 years and 45% of the learners were female.

### Academic Motivation Scale (AMS)

Vallerand and colleagues have added further theoretical concepts to the model of Deci and Ryan (1985) as well as adjusting the model for different contexts as SDT was primarily developed to measure motivation among children. Individual motivation was measured by the Academic Motivation Scale (AMS), which was developed by Vallerand et al. (1992) for college/university learners and measures the contextual motivation for education. The instrument consists of 28 items, in all of which learners respond to the question stem "Why are you going to college?". There are seven subscales on the AMS, of which three belong to the intrinsic motivation scale, three to the extrinsic motivation scale and one for amotivation. The response-rate on AMS-questionnaire among the summer course participants was 93%. The Cronbach alpha reliability for the seven scales ranged from .760 to .856. The 82 students who participated in our setting were unaware of the scores of the AMS and those of their peers with whom they worked and learned together in their virtual team.

### Statistical analyses

*Cluster analysis*

The 82 students in the experiment are part of the inflow of 765 international freshmen. Motivational profiles were determined of all international students by applying k-means cluster analysis to subscale scores of the AMS-instrument. It was decided to base motivational profiles on the complete sample of international freshmen. First of all, this will lead to a more stable outcome of the cluster analysis. Second, in this way we are able to express the motivational patterns found amongst participants relative to the motivational profiles present amongst all international students. Since participation in the experiment is voluntarily, motivation scores of participating freshmen might be different from motivation scores of all freshmen. In this situation, profiles found amongst all international students were regarded as more relevant benchmarks than profiles found in the restricted group of participants of the experiment. It was found that a three cluster solution provides the best fit for different motivation profiles present in these freshmen. Afterwards, data on cluster membership of all participants of the virtual networks were combined with individual data resulting from the social network analysis. The interrelationships between all measures were assessed through standard T-tests analyses using SPSS 15.0.1.

*Positioning of individuals within social network using SNA*

Two SNA measures were used, namely ego network density, which measures to how many other learners a learner is directly connected, and Freeman's degree of Centrality, which measures whether learners were central in the social network or not (Freeman, 2000; Wassermann & Faust, 1994). Main indicator for this study is the relative position of each learner within the social network, derived by UCINET version 6.158. In order to assess whether learners with different motivational orientations connect equally to each of the clusters, we will use the (absolute/relative) number of send and received messages per learner to members in each of the (send to own/outside) clusters as a measurement for equality of interaction between clusters. An innovative feature of this study is that by combining the results of the SNA and cluster analysis, we were able to distinguish interaction patterns amongst individual learners based upon their motivation profile.

## Results

In order to test hypothesis 1, the average number of connections in the cohort of online summer course participants is compared. On average, a learner has 6.43 (SD = 4.03) connections to other learners and there are substantial differences amongst individual learners with respect to the number of connections as assessed by a Chi-Square test ($\chi 2$ (df = 76) 159.46, $p < .001$). Furthermore, significant differences are found using a Chi-Square test in each of the six virtual networks with the exception of network 3. In other words, in contrast to random graph theory the social networks in our setting do not evolve to a random network with an approximately equal amount of connections per learner, with the exception of network 3. Furthermore, some learners are more central than other learners in the network, as is illustrated by the large standard deviation of the Freeman's degree of centrality (M = 26.60, SD = 24.29), as well as by the Chi-Square test for all participants ($\chi 2$ (df = 80) 1772.74, $p < .001$) and the Chi-Square test for participants within each of the networks. As a result, we need to reject hypothesis 1 that social networks develop and evolve in accordance to the model of the random graph theory for five out of six of our networks.

In order to test hypotheses 2 and 3 and to investigate whether the motivation profile of a learner has an influence on the formation of links to other learners within the social network, a K-means cluster analysis is applied to obtain three different profiles for motivation, which are further labeled according to the final cluster center position (see Table 1). The three motivation profiles are: cluster 1: low intrinsic motivation (Low In), high extrinsic motivation (High Ex); cluster 2: medium intrinsic motivation (Med In), low to medium extrinsic motivation (Med Ex); cluster 3: high intrinsic motivation (High In), high extrinsic motivation (High Ex).

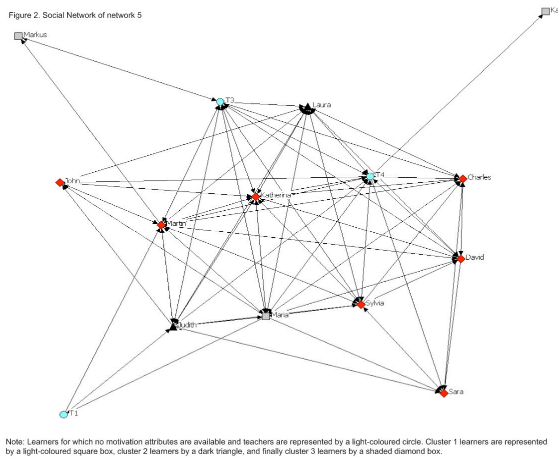
Table 1 Means and standard deviation of classification measures per cluster (K-means)

| | Cluster 1 Low In, High Ex (N=182) | Cluster 2 Med In, Med Ex (N=152) | Cluster 3 High In, High Ex (N=415) |
|---|---|---|---|
| Intrinsic motivation to know (IMTK) | 4.68 (0.94) | 5.38 (1.02) | 6.06 (1.10) |
| Intrinsic motivation to accomplish (IMTA) | 3.95 (0.89) | 4.09 (0.89) | 5.42 (1.06) |
| Intrinsic motivation to experience stimulation (IMES) | 3.17 (0.95) | 3.81 (0.99) | 4.92 (1.18) |
| Identified regulation (EMID) | 6.04 (1.00) | 5.58 (1.20) | 6.48 (1.03) |
| Introjected regulation (EMIN) | 4.61 (1.14) | 3.24 (1.23) | 5.35 (1.22) |
| External regulation (EMIR) | 6.05 (1.03) | 4.52 (1.43) | 6.12 (1.23) |
| Amotivation (AMOT) | 1.44 (0.73) | 1.40 (0.73) | 1.32 (0.62) |

As a third step, the cluster memberships are added as learner attributes to the social networks of each of the six virtual networks. Based upon the division of motivational profiles, network 5 (Figure 2) and network 6 (Figure 3) can be categorised as prototypical networks. Learners for which no motivation attributes are available and teachers are represented by a light-coloured circle, while cluster 1 learners (Low In, High In) are represented by a light-coloured square box, cluster 2 learners (Med In, Med Ex) by a dark triangle, and finally cluster 3 learners (High In, High Ex) by a shaded diamond box[1]. In this way, we are able to visualise the position of each learner

[1] The names of the participants are replaced by fictitious names in order to guarantee privacy of the participants.

in the network as well as to whom each learner is connected to depending on his/her motivational profile. When looking at the three motivation profiles, it appears that learners with high intrinsic motivation are situated closely together. For example, in network 6 most of the connections of Veronica and Jonas (cluster 3) are to learners with the same cluster membership. Learners with low and medium motivation are positioned mostly on the outer fringe of the network and are mainly connected to highly intrinsically motivated learners. Furthermore, learners within cluster 1 (Kathi and Markus of network 5; Paul and Bart of network 6) and learners within cluster 2 (Judith and Laura; Elena, Christina and Bernard) are not well connected to other learners with the same motivation profile. In fact, most cluster 1 and 2 learners are only indirectly linked to each other through cluster 3 learners. For example, in network 6 Bart can only be linked to Paul via Jonas or Caroline. In sum, our learners differ with respect to the number of ties as well as with respect to the position in the network. Furthermore, we find that the position of learners in a social network depends on the type of motivation. Cluster 3 learners form the center core of the network, while the other learners are mostly situated on the outer fringe.



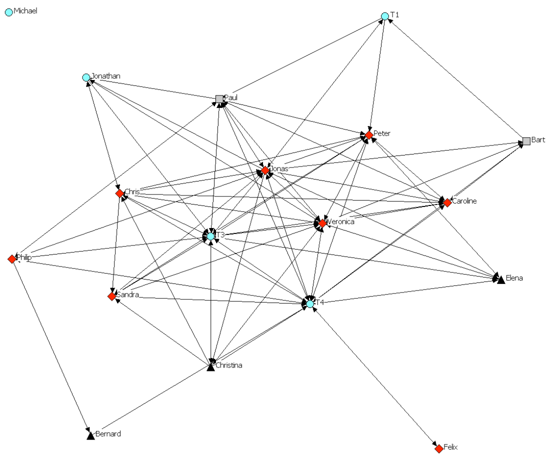




Figure 2. Social Network of network 5 Figure 3. Social Network of network 6

Table 2 Interaction among learners per cluster corrected by relative cluster size

| | Cluster 1 Low In, High Ex (N=15) | Cluster 2 Med In, Med Ex (N=12) | Cluster 3 High In, High Ex (N=50) | t-test difference |
|---|---|---|---|---|
| Sent total | 1.60 (1.50) | 2.30 (2.45) | 2.74 (2.74) | 1.950† |
| Sent to own cluster | 0.62 (0.67) | 1.22 (1.54) | 1.70 (1.71) | 2.790** |
| Sent outside own cluster | 0.97 (0.98) | 1.07 (0.95) | 1.04 (1.18) | 0.518 |
| Sent difference | -0.35 (0.76) | 0.15 (0.75) | 0.66 (1.08) | 3.80*** |
| Received total | 1.67 (1.64) | 2.32 (2.23) | 3.25 (2.57) | 2.824** |
| Received from own cluster | 0.62 (0.72) | 1.17 (1.34) | 1.84 (1.68) | 3.356*** |
| Received from outside own cluster | 1.04 (1.09) | 1.15 (1.07) | 1.41 (1.17) | 1.660 |
| Received difference | -0.42 (0.86) | 0.02 (0.95) | 0.42 (1.10) | 3.033** |

Note: Independent sample T-test (2-sided) (Cluster 1 + 2 vs. Cluster 3)
† Coefficient is significant at the 0.10 level (2-tailed).
* Coefficient is significant at the 0.05 level (2-tailed).
** Coefficient is significant at the 0.01 level (2-tailed).
***Coefficient is significant at the 0.001 level (2-tailed).

In Table 2, the relative interactions within and between clusters are illustrated, whereby we correct for the total number of each of the three profiles of motivation within a virtual network. Both cluster 1 and 2 differ significantly from cluster 3 using an independent sample T-test with the exception of sent outside own cluster and received from outside own cluster. For all cluster 1 learners in the six networks, this implies that on average 0.62 messages are sent to each of the cluster 1 learners. At the same time, on average 0.97 messages are sent by cluster 1 learners to each learner outside their own cluster. That is, cluster 1 learners send on average 56% more messages outside their cluster and this difference is significant at 10% ($T = -1.768$, $p < 0.10$) in a paired-samples T-test. At the same time, cluster 1 learners receive 68% more external messages from outside their cluster than from inside their cluster and this difference is again significant at 10% ($T = -1.883$, $p < 0.10$) in a paired-samples T-test. Therefore, both sent to and received from measures indicate that cluster 1 learners are mainly focussed on communication with learners outside their own cluster, implying that the motivation profile has an influence on whom cluster 1 learners are connected to. In other words, we find support for hypothesis 2 that extrinsically motivated learners are more likely to interact with intrinsically motivated learners than with extrinsically motivated learners.

Cluster 2 learners (medium intrinsic, low to medium extrinsic motivation) send about an equal amount of messages to both within and outside their cluster. At the same time, they receive an equal amount of messages from within as well as outside their cluster. This implies that cluster 2 learners do not distinguish with whom they communicate. Thus, cluster 2 learners are connected to other learners within the social network as predicted by random graph theory (Barabási, 2002; Barabási & Albert, 1999; Erdős & Rényi, 1960).

Finally, cluster 3 learners contribute most actively to discourse in absolute and relative numbers. More messages are contributed to learners within the same cluster, namely 1.70 messages per learner in cluster 3. In contrast, only 1.04 messages are sent to each learner outside their own cluster. In other words, cluster 3 learners are almost 40% more likely to send a message to their own cluster and this difference is statistically significant at 1% ($T = 4.326$, $p < 0.01$) in a paired samples T-test. In addition, the majority of the messages received by learners in cluster 3 originate from their own cluster ($T = 2.748$, $p < 0.05$). If we subtract the average number of contributions sent to external clusters (1.04) from those received from external clusters (1.40), we find that the communication of cluster 1 and 2 members is more strongly directed to cluster 3 members than vice-versa, and this difference is significant ($T = -3.879$, $p < 0.01$) in a paired-samples T-test. Hence, the stronger extrinsically motivated learners, and the learners with a less outspoken motivational profile, are connecting primarily to the intrinsically motivated learners, which supports hypothesis 2. Last but not least, intrinsically motivated learners are the most active contributors to discourse, but, in agreement with hypothesis 3, are contributing mostly with learners having similar motivational profile.

## Discussion

The results of the present study indicate that in our settings learners connect to other learners in their virtual network depending on their motivation profile. We find evidence that learners with high intrinsic motivation receive a relatively large amount of contributions from learners with other motivational profiles. At the same time, intrinsically motivated learners themselves are focussing more on discourse with other intrinsically motivated learners. These findings indicate that in distance learning settings interaction patterns amongst participants and evolutions of social networks of virtual networks do not develop randomly. In fact, we find that highly extrinsically motivated learners are more likely to connect to intrinsically motivated learners than vice versa. A new feature is that we are able to link the position of the learner in the virtual network to his/her motivational profile. Most extrinsically motivated learners seem to be stronger connected to intrinsically motivated learners than vice versa.

These findings might have important consequences for instructional designer and online teachers (Mishra & Koehler, 2006; Wang, 2009) as we find support of the idea that in distance learning settings learners prefer to interact with learners who are highly intrinsically motivated (Martens et al., 2004; Rienties et al., 2009). This implies that learners strong in intrinsic motivation, who due to the nature of distance learning already have an advantage over other learners (Rienties et al., 2009), will in the duration of the course be further stimulated by extrinsically motivated learners as well as other intrinsically motivated learners that are keen to link to them. By receiving more contributions from others to initiated discourse (in particular from intrinsically motivated learners), they can exchange more knowledge and receive more feedback than learners with low intrinsic motivation who receive little contributions from others. In a way, it seems like a self-fulfilling prophesy: active contributors to discourse receive further encouragements from others to continue, while these active contributors at the same time interact mostly with other active contributors rather than learners on the outer fringe of the network. Therefore, intrinsically motivated learners appear “well-suited” for our distance learning setting and continuously receive acknowledgements from other learners (Martens et al., 2004). Given that many educational psychologists have found that learners who are actively co-constructing knowledge eventually have a deeper learning experience (Hmelo-Silver, 2004; Järvelä et al., 2008; Van den Bossche et al., 2006), receiving a lack of reply on contributions might have a negative impact on learning for extrinsically motivated learners. As a result, extrinsically motivated learners receive less feedback and stimuli from others, which might further decrease their integration within the virtual network.

The role of the teacher in designing a challenging and interactive learning environment (Mishra & Koehler, 2006; Wang, 2009) for all types of learners seems to be a prerequisite for interactive learning for all learners. Furthermore, a helpful tool for teachers to understand the complex dynamics of social interaction is to use the insights from motivational science to enhance learning. We suggest that teachers ask their students to fill in a motivation questionnaire before the beginning of the course. This can for example be the Academic Motivation Scale developed by Vallerand and colleagues (1992) or the Quality of Working in Groups Instruments developed by Boekaerts and Minnaert (2006). The results can be used to assess what type of motivated learners teachers have in their course and to actively stimulate students who are less active in discourse.

## Limitations

The results of this study were based on a k-means cluster analysis on learner self-scores for a questionnaire on academic motivation, which was afterwards linked to the social network of each virtual network using Social Network Analysis. This can be viewed as a potential limitation to this study as a self-reported measurement of academic motivation was used with obvious limitations. However, the patterns of interaction among the three identified motivational profiles follow the anticipated direction. In addition, research by Vallerand and colleagues has found that the AMS instrument is a robust predictor of learning outcomes and academic performance. As a second limitation, the long-term consequences on learning outcomes have not been demonstrated. However, our longitudinal analysis of learning outcomes among summer course participants indicate that active summer course participants outperform others in the first year of their bachelor programme (Rienties, Tempelaar, Dijkstra, Rehm, & Gijselaers, 2008). A third limitation of this study is that no measures were taken to prevent self-selection in the summer course programme. In our setting, which matches the practice teachers in online settings are confronted with (i.e. networks with a mix of various types of motivated learners), we did not balance networks based on a pre-determined mix of motivational types. We established that the proportion of cluster 3 learners amongst summer course participants is indeed somewhat higher than the proportion in all freshmen, yet cluster 1 and cluster 2 learners are not statistically significantly underrepresented in our subsample. So selection effects, if present, are of limited size.

### Future Research and Implications for Education

Based on our findings, we will redesign the learning environment to capitalise on the merits of social interaction, peer-support and planning of learning processes. By increasing social presence in our virtual learning environment by using Web 2.0 tools like wiki's and web-videoconference, we hope to increase the relatedness among learners, which has shown to increase the internalisation of motivation regulation (Ryan & Deci, 2000). Socio-emotional support is an important factor in relational development of networks. In particular in CMC environments, socio-emotional communication is not an automatic artefact. These findings are relevant for teachers, managers, admission officers and schedulers as the results imply motivational orientation has a moderately strong influence on the type of discourse and position within the social network. Social Network Analysis tools can be used to assess who is contributing actively to discourse and can be used as a tool for teachers to identify learners on the outer ring of the social network. Appropriate strategies to deal with various types of motivation should be designed to assist each type of learner.

## References


Barabási, A.-L. (2002). *Linked: The New Science of Networks*. New York: Perseus Publishing.

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science, 286*(5439), 509-512.

Beers, P., Boshuizen, H., Kirschner, P. A., & Gijselaers, W. H. (2005). Computer support for knowledge Construction in collaborative learning environment. *Computers in Human Behaviour, 21*, 623-643.

Beers, P., Boshuizen, H., Kirschner, P. A., & Gijselaers, W. H. (2007). The analysis of negotiation of common ground in CSCL. *Learning and Instruction, 17*(4), 427-435.

Boekaerts, M., & Minnaert, A. (2006). Affective and Motivational Outcomes of Working in Collaborative Groups. *Educational Psychology: An International Journal of Experimental Educational Psychology, 26*(2), 187 - 208.

Bromme, R., Hesse, F. W., & Spada, H. (Eds.). (2005). *Barriers and Biases in Computer-Mediated knowledge communication and how they may be overcome* (5 ed.). New York: Springer.

Caspi, A., Chajut, E., Saporta, K., & Beyth-Marom, R. (2006). The influence of personality on social participation in learning environments. *Learning and Individual Differences, 16*(2), 129-144.

Caspi, A., Gorsky, P., & Chajut, E. (2003). The influence of group size on nonmandatory asynchronous instructional discussion groups. *The Internet and Higher Education, 6*(3), 227-240.

De Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Online teaching in networked learning communities: A multi-method approach to studying the role of the teacher. *Instructional Science, 35*(3), 257-286.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum.

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5*, 17–61.

Freeman, L. (2000). Visualizing Social Networks. *Journal of Social Structure, 1*(1).

Giesbers, B., Rienties, B., Gijselaers, W. H., Segers, M., & Tempelaar, D. T. (2009). Social presence, web-videoconferencing and learning in virtual teams. *Industry and Higher Education, 23*(4), 301-310.

Hmelo-Silver, C. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review, 16*(3), 235-266.

Hurme, T., Palonen, T., & Järvelä, S. (2007). Metacognition in joint discussions: an analysis of the patterns of interaction and the metacognitive content of the networked discussions in mathematics. *Metacognition and Learning, 1*(2), 181-200.

Järvelä, S., Järvenoja, H., & Veermans, M. (2008). Understanding the dynamics of motivation in socially shared learning. *International Journal of Educational Research, 47*(2), 122-135.
Jonassen, D. H., & Kwon, H. (2001). Communication patterns in computer mediated versus face-to-face group problem solving. *Educational Technology Research and Development, 49*(1), 35-51.
Kirschner, P. A., Beers, P., Boshuizen, H., & Gijselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior, 24*(2), 403-420.
Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2003). What Makes a Student Group Successful? Student-Student and Student-Teacher Interaction in a Problem-Based Learning Environment. *Learning Environments Research, 6*(1), 59-76.
Martens, R., Gulikers, J., & Bastiaens, T. (2004). The impact of intrinsic motivation on e-learning in authentic computer tasks. *Journal of Computer Assisted Learning, 20*, 368-376.
Martinez, A., Dimitriadis, Y., Rubia, B., Gomez, E., & de la Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education, 41*(4), 353-368.
Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record, 108*(6), 1017-1054.
Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review, 45*(2), 167-256.
Resta, P., & Laferrière, T. (2007). Technology in Support of Collaborative Learning. *Educational Psychology Review, 19*(1), 65-83.
Rienties, B., Tempelaar, D. T., Dijkstra, J., Rehm, M., & Gijselaers, W. H. (2008). Longitudinal study of online remedial education effects. In N. P. Barsky, M. Clements, J. Ravn & K. Smith (Eds.), *Advances in Business Education and Training 1: The Power of Technology for Learning* (pp. 43-59). Dordrecht: Springer.
Rienties, B., Tempelaar, D. T., Van den Bossche, P., Gijselaers, W. H., & Segers, M. (2009). The role of academic motivation in Computer-Supported Collaborative Learning. *Computers in Human Behavior, 25*(6), 1195-1206.
Rienties, B., Tempelaar, D. T., Waterval, D., Rehm, M., & Gijselaers, W. H. (2006). Remedial online teaching on a summer course. *Industry and Higher Education, 20*(5), 327-336.
Russo, T. C., & Koesten, J. (2005). Prestige, Centrality, and Learning: A Social Network Analysis of an Online Class. *Communication Education, 54*(3), 254-261.
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology, 25*(1), 54-67.
Schellens, T., & Valcke, M. (2005). Collaborative learning in asynchronous discussion groups: What about the impact on cognitive processing? *Computers in Human Behavior, 21*(6), 957-975.
Tu, C., & McIsaac, S. M. (2002). The Relationship of Social Presence and Interaction in Online Classes. *American Journal of Distance Education, 16*(3), 131-150.
Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C., & Vallières, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52*, 1003–1017.
Van den Bossche, P., Gijselaers, W. H., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments. Team learning beliefs & behaviour. *Small Group Research, 37*, 490-521.
Veermans, M., & Lallimo, J. (2007). Multimethod approach for analyzing students' motivational profiles and their participation in virtual collaboration. *Qwerty, 1*, 17-30.
Wang, Q. (2009). Designing a web-based constructivist learning environment. *Interactive Learning Environments, 17*(1), 1 - 13.
Wassermann, S., & Faust, K. (1994). *Social Network Analysis: methods and applications.* Cambridge: Cambrdige University Press.
Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71-95.
Yang, C.-C., Tsai, I. C., Kim, B., Cho, M.-H., & Laffey, J. M. (2006). Exploring the relationships between students' academic motivation and social ability in online learning environments. *The Internet and Higher Education, 9*(4), 277-286.

**Acknowledgments**
The authors would like to thank the EU LLP programme funding the S.T.E.P. project, which enabled this research project. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

# Investigating pre-service elementary teachers' epistemologies when talking about science, enacting science and reflecting on their enactment

Louca, T. L., Tzialli, D., Department of Education Sciences, European University,
6, Diogenous Str., Engomi, 1516 Lefkosia, Cyprus,
Louca.L@cytanet.com.cy, D.Tzialli@euc.ac.cy

Zacharia, C. Z., Department of Educational Sciences, University of Cyprus,
P.O. Box 20537, Lefkosia 1678, Cyprus, zach@ucy.ac.cy

**Abstract:** We described and compared 94 pre-service elementary teachers' epistemologies during three different activities: one semi-structured interview, an asynchronous on-line discussion about a physics problem and their reflection on the discussion of the second activity. Using discourse-based analysis, we analyzed the data in terms of the teachers' underlying epistemologies and findings revealed significant differences across the three activities. This suggests that (a) teachers' epistemologies might be better understood as finer grained cognitive resources whose activation is sensitive to the context, unlike most research which views them as coherent and stable cognitive structures, and that (b) the research community is far from settling the debate as to what particular approaches should be used to assess or study personal epistemologies. Depending on the context and the manner of investigation, students and teachers may "show" different epistemological understanding.

## Introduction

Current science education standards (NRC, 1996; 2007) argue that inquiry should be a central strategy of science instruction, partly because that inquiry can be a dynamic way for helping students to develop sophisticated understanding of the nature of science (NOS) or sophisticated scientific epistemologies. Epistemology is a term used differently by philosophers and psychologists (Sandoval, 2005). For the purposes of this paper we take scientific epistemologies to mean the individual's understanding of or views about the nature and characteristics of scientific knowledge and its construction, including the sources of scientific knowledge, its truth value and what constitutes scientifically appropriate warrants, and science as a way of knowing and its values inherited in the development of the scientific knowledge (Abd-El-Khalick & Lederman, 2000; Abd-El-Khalick et al., 1998; Lederman, 1992). The development of sophisticated epistemologies has been a central component of scientific literacy for some years (AAAS, 1993; NRC, 1996).

However, this agenda has been slow to become established in instructional practice for a number of reasons. First, despite a wide consensus for the importance of developing epistemological awareness in science among learners, the education community has yet to agree on precisely how this awareness is developed, how epistemological knowledge is developed and in what form (Louca et al, 2004). Most research on personal epistemologies has assumed they consist of developmental stages or beliefs and thus can be explicitly taught and developed (van Driel et al., 2001; Olson, 2007). Others have advocated a structure of epistemologies that consists of units of cognitive structure at a finer grain size than stages or beliefs, suggesting that it is not really a matter of developing new knowledge constructs, but a matter of organizing a network of existing epistemological resources (Louca et al, 2004; Hammer & Elby, 2002). Second, there is no consensus regarding the methodological approaches used to study students' epistemologies (Sandoval, 2005). One approach has argued that students' epistemologies can be directly investigated, e.g., via interviews (Van Driel et al., 2001). A second approach has suggested that students' personal epistemologies are manifested through practice, and, thus, practice is the only way of studying them (Sandoval, 2005; Richardson, 1996). These ambiguities have resulted in a rather fragmented picture of students' epistemologies, with often contradicting findings and frequently students' practices of inquiry appear to share much with scientific practice but their expressed epistemologies are naïve (Sandoval, 2005).

Our purpose in this paper was to investigate 94 pre-service elementary teachers' epistemologies while engaged in three different activities. The differences among these activities refer to the context in which we investigated their expressed or enacted epistemologies. Firstly, participants were prompted to talk about their personal epistemologies during an individual interview. Secondly, they were asked to "enact" science, by talking about a physics topic on relative motion through an online discussion (without specifically talking about their personal epistemologies). By enactment, we refer to the participants' efforts to reach a consensus about a problem related to relative motion. In some cases this involved students trying out experiments and reporting them to the group or debating various points of view. Thirdly, they were asked to reflect on their enactment of science during the second activity by debating whether their discussion was scientific or not – thus, prompting them to reflect on the NOS underlying their discussion of relative motion.

## Theoretical framework

### Current views of NOS

Traditionally, science has been presented to students as a rigid body of facts, theories, and rules to be memorized and practiced, rather than a way of knowing about natural phenomena (van Driel et al., 2001). The learning of correct answers, memorizing of information, rote learning, and reading were central components of the exploration of questions, critical thought, understanding in context, argument, and doing science. In this view, teaching science assumed that an already developed body of knowledge, one generally accepted by the scientific community, could be transmitted to students through passive instructional means (Tobin, Tippins, & Gallard, 1994).

To an increasing extent this approach has become the subject of criticism among policy makers, teachers, educators, and researchers (e.g., Tobin et al., 1994). This approach has been related to the decreasing popularity of science among students (van Driel et al., 2001). Furthermore, research on students' scientific epistemologies has demonstrated that students exposed to this approach often end up with a poor understanding of scientific concepts (van Driel et al., 2001). Moreover, science education in its traditional form fails to adequately prepare future citizens to understand science and technology issues in a rapidly evolving society (Millar & Osborne, 1998).

Recently, modern views of NOS suggest that scientific knowledge is (a) tentative (that is, subject to change), (b) empirically based (that is, testable), (c) subjective (and thus, not absolute), (d) creative, (e) socially and culturally embedded, (f) unified, (g) amoral, (h) parsimonious (scientific knowledge attempts to achieve a simplicity of explanation as opposed to complexity), and (i) differentiates between observation and inference (Abd-El-Khalick, Bell, & Lederman, 1998). In this sense, scientific knowledge is based on observations from the natural world and reflects human attempts to impose order on the understanding of nature and explain the mechanism underlying the physical phenomena.

Reform documents (e.g., AAAS, 1990; NRC, 1996) and researchers (e.g., Lederman, 1992; McComas & Olson, 1998) have advocated that there are several elements of NOS that should be communicated to students in order for them to become scientifically literate. Thus, aspects of NOS are crucial to guide the current assessment of individuals' understandings of NOS, because they are not contentious, they are developmentally appropriate for K–12 students, they are learnable by K–12 students as indicated by empirical research, and they are arguably important for all high school graduates to know. Consequently, the objective of helping students (and teachers for that manner) to develop understandings of NOS is one of the most common objectives for science education (Abd-El-Khalick, Bell, & Lederman, 1998).

### Views about the nature of scientific epistemologies

Disagreements about the tenets of NOS exist at a general level among scientists, science educators, historians and philosophers of science. Despite these disagreements, there is a widely shared view which suggests that individuals' epistemologies have a developmental aspect and that these are learnable throughout K-12 (Abd-El-Khalick & Lederman, 2000). This perspective presents the development of epistemologies as developmental stages somewhat resembling cognitive development as conceptualized e.g., by Piaget (1967). For instance, in the seven-stage scheme of King and Kitchener (1994), children initially view knowledge as comparatively certain and gained from authority or direct observation. From there they progress to relativist stages in which they view knowledge as constructed and different viewpoints as valid. Finally, some reach expert stages in which they see knowledge as constructed yet subject to scrutiny, judgment, and synthesis. Invoking more recent developmental theorists such as Fischer's (1980), King and Kitchener (2004) adopt a "complex stage theory" in which a typical subject's epistemology is like a wave spread over two stages.

A second perspective suggests that students' epistemologies are not a unified whole. Following a growing set of studies that suggest that students' epistemological views of science are not stable coherent frameworks, but inconsistent, fragmented and possibly unstable beliefs, this perspective deals separately with each different dimension of NOS, such as, structure, certainty and source of knowledge (Hofer & Pintrich, 1997). It suggests that learners' epistemologies along each dimension are assumed to consist of semi-independent beliefs, implying that a student could hold sophisticated views about the structure of e.g., physics knowledge, seeing it as a hierarchy of concepts rather than a collection of equations, while also holding naïve beliefs about the certainty of that knowledge, viewing new theories as fixed and absolute. In Leach et al.'s (2000) study, students' responses to decontextualized and contextualized open-ended survey items that asked them to reason about the relation between theory and data were found to be inconsistent across the two contexts. Sandoval and Morrison (2003) interviewed a sample of high school students before and after a month-long intervention and found that both individual students' responses to different questions reflected different epistemological levels, and that student responses were not stable across interviews (nor predictable). According to this perspective, views about NOS consist largely of comparatively stable, robust cognitive structures corresponding to articulate, declarative knowledge. These epistemologies are taken to be the units of views about NOS (Schommer-Aikins, 2004). Some researchers view epistemological beliefs as comparatively global (Schommer, 1990), while others have investigated how epistemological beliefs vary by discipline, e.g., in chemistry vs. psychology (Hofer, 2002a). What they all agree upon, however, is that epistemologies consist largely of comparatively stable, robust cognitive structures corresponding to articulate, declarative knowledge.

Hammer and colleagues (Hammer & Elby, 2002; Louca et al, 2004) have proposed a third perspective, suggesting that personal epistemologies are made up units of cognitive structure at a finer grain size than stages or beliefs, which they call resources. Rather than attribute to individuals any general epistemological theories or beliefs, they understand them to have a range of cognitive resources for understanding scientific knowledge and its nature. They suggest that the difference between naïve and expert epistemologies lies not just in the content (views), but also in the form of the relevant cognitive elements. Other studies show that students often hold inconsistent epistemologies which emerge in different contexts (Hammer, 1994; Roth & Roychoudhury, 1994; Solomon, Duveen, & Scott, 1994). There is also some evidence for fragmented epistemologies from those studies that have been unable to assign large portions of students to a single epistemological "type" (Khishfe & Abd-El-Khalick, 2002; Linn & Songer, 1993). In this sense, a teacher's professed epistemology, for example, her stated views about knowledge and learning, may possibly differ substantially from her enacted epistemology, the views about knowledge and learning an observer would infer from her classroom behavior (Hofer, 2002b; Tobin & McRobbie, 1997), suggesting that it could be a matter of different resources being activated in different contexts – that of an interview and of teaching contexts (e.g., Louca et al., 2004; Leach et. al, 2000; Roth & Roychoudhury, 1994; diSessa, Elby, & Hammer, 2002).

## Novice teachers' epistemologies

A large number of studies have indicated that teachers' epistemologies influence their teaching and their students' learning to a great extent (AAAS, 1993; NRC, 1996). As a result, researchers have investigated teachers' epistemologies, in addition to developing programs that seek to improve them (e.g., Abell & Smith, 1994; Palmquist & Finley, 1997; Abell et al., 2001). Many of these studies were consistent in showing that teachers (both in-service and pre-service) possessed naive scientific epistemologies (e.g., Abd-El-Khalick & BouJaoude, 1997).

Moreover a number of these studies has provided insights about how novice teachers (pre-service and early-career teachers) encounter teaching and their ability to reflect on practical experience (Penso, Shoman, & Shiloah, 2001), how science teachers' views about science and science teaching influence their classroom practice (Brickhouse & Bodner, 1992) and how induction programs are essential in addressing the pedagogical and content needs of science teachers (Luft, Roehrig, & Patterson, 2003). Other studies have documented the nature and persistence of pre-service teachers' epistemologies. For instance, many pre-service teachers think of teaching as passing on knowledge, and learning as absorbing and memorizing knowledge (e.g., Calderhead & Robson, 1991). When they imagine themselves teaching, pre-service teachers often picture themselves standing in front of a group of attentive students presenting information, going over problems and giving explanations. A significant proportion of teachers believe that scientific knowledge is not tentative, and hold a positivistic, idealistic view of science (Lederman, 1992).

## The relationship between teachers' views of NOS and classroom practice

Research concerning improving teachers' epistemologies has followed the assumption that teachers' epistemologies directly transfer into their classroom practices (Lederman, 1992). In other words, it is assumed that studying teacher practices can help understand their epistemologies (Sandoval, 2005), and that improving teachers' epistemologies is sufficient for promoting "effective" NOS instruction in science (Lederman, 1992).

However, through a range of studies, it is currently understood that the relationship between teachers' epistemologies and their classroom practice is more complex than originally assumed. Several factors have been identified to mediate and constrain the translation of teachers' epistemologies into practice. These factors include pressure to cover content (Abd-El-Khalick et al., 1988; Hodson 1993), classroom management and organizational principles (Hodson, 1993), concerns for student abilities and motivation (Brickhouse & Bodner 1992; Lederman, 1999), institutional constraints (Brickhouse & Bodner, 1992), teaching experience (Lederman, 1999), discomfort with understandings of NOS, and the lack of resources and experiences for assessing student epistemologies (Abd-El-Khalick et al., 1998).

Research on the translation of teachers' epistemologies into classroom practice indicates that even though these can be thought of as a necessary condition for promoting the development of students' NOS understanding, these epistemologies should not be considered sufficient (Lederman 1992). This may suggest that research efforts should concentrate beyond simply identifying teachers' professed epistemologies, to investigating their translation into enacting (or practical) epistemologies as they appear during everyday teaching. By teachers' practical epistemology we refer to the set of epistemologies about ones' own knowledge production in school science, including what knowledge is, the methods through which knowledge can be produced, and the criteria for evaluating knowledge claims, which are reflected in the epistemic decisions people make during the construction and evaluation of scientific knowledge (Sandoval, 2005).

All these issues raise three questions about teachers' scientific epistemologies and their development. First, the nature of epistemologies remains somewhat unclear; whether epistemologies are organized in coherent frameworks, or are fragmented beliefs or finer-grain resources. Second, the specific epistemologies that guide teachers' practices are largely unknown. Third, the relation of these practical epistemologies and teachers' expressed epistemologies are not well articulated. Given the discrepancy between what teachers seem able to do and the difficulty they have in articulating epistemological aspects of formal science, it seems more likely that epistemological beliefs are contextualized rather than coherent frameworks (Hammer & Elby, 2002).

Our purpose in this paper is to contribute specifically to these needs: to investigate teachers' expressed and practical epistemologies through multiple data sources. The differences among these data sources refer to the context in which we investigated their expressed or enacted epistemologies. Thus, we compare 94 pre-service teachers' expressed epistemologies as revealed through an interview, their practical epistemologies underlying a discussion about relative motion, and their epistemologies underlying a reflective discussion about whether their science enactment in the relative motion discussion was scientific or not.

## Methodology

The present study was interpretive in nature and focused on the epistemologies that participants held, expressed or underlay their actions in the three aforementioned tasks. Ninety-four pre-service elementary teachers participated in this study. All participants were enrolled in a science methods course in the same semester (separated in two groups), taught by the first author. The data collection took place at the very beginning of the course (during the first three weeks), prior to any intervention and discussion about NOS characteristics.

Data sources consisted of three different activities. For the first activity we randomly selected 47 out of the 94 pre-service teachers of the study, whom we interviewed individually about their views concerning NOS. All interviews were conducted by the second author, using a semi-structured protocol which focused on three areas, namely, demographics, views about science and about teaching science. The first part of the interview focused on identifying pre-service teachers' experiences with science from both pre-university and university studies. The second part of the interview aimed at identifying pre-service teachers' epistemologies (for instance, we asked them what is Science for them; under what conditions is something scientific; and what does the term "experiment" mean to them). The third part of the interview focused on investigating pre-service teachers' views about science teaching. For instance, we asked them what they thought the goal of Science education is; about the role of a teacher during a science lesson; and what they thought the components of a "good" science lesson are. Finally, interviewees watched a short video clip from a science lesson and were asked to comment about the ways they might respond to students' ideas and reasoning, had they been the teacher during this lesson.

For the second activity, all 94 teachers had an online, asynchronous threaded discussion about relative motion. We randomly assigned pre-service teachers into groups of 10 or 11 participants and provided them with a question on relative motion: A person is running, holding a set of keys in her hand, which she holds still next to her body. What would happen to the keys, if she let go, while still running? We asked each pre-service teacher to have at least one posting of her own answer to the question, respond to at least three other group-mates' answers, respond to all the comments she received for her own postings, including her answer to the initial question, in an effort to reach consensus about the answer to the question, after discussing this with the rest of the members of her group. The pre-service teachers had a week in which to reach a consensus. A total of 341 posts were made by the 94 teachers which reflected an average of 80 words per post, and all groups of students were close to, or reached consensus. The third activity was also an online, asynchronous threaded discussion, carried out by the same groups of pre-service teachers as above. This occurred immediately after the second activity and also lasted for a week. In this activity, pre-service teachers had to reflect on the discussions they had during the second activity, in an attempt to discover whether that discussion was scientific or not. A total of 359 posts were made by the 94 teachers, which reflected an average of 105 words per post.

All interviews were transcribed and along with all threaded discussions served as the primary sources of data. We analyzed all data using a discourse-based analysis focusing on three particular aspects of NOS: (a) what is science (b) what is the result of science and (c) what is scientific knowledge? Categories for this analysis were drawn from the literature we have described in the theoretical framework of the paper. Some categories identified in the literature were not observed in our data and thus are not part of our coding scheme. The final categories that we used for each aspect of NOS were organized in an ordinal scale based on their sophistication. Each teacher's postings for a particular discussion were analyzed separately and coded in terms of the underlying ideas about the three aspects of NOS. The same was done with each participant's interview. In all cases, each teacher's postings or interview was assigned with only one code for each of the three aspects of the

NOS. All codings were carried out independently by the first and the second author (Cohen's Kappa=0,84), and differences were resolved through discussion.

We used descriptive statistics to compare teachers' views about the result of science, science as a process and the scientific knowledge from the three different contexts. Lastly, the non-parametric Wilcoxon Signed-Rank Test was applied to test for significant differences in teachers' views among the three data sources (contexts). The test was applied only to the 47 pre-service teachers from whom we had data from all three contexts.

## Findings

Findings revealed considerable differences among teachers' epistemologies about NOS in the three contexts (see Table 1). Activity 1 will be referred to as "interview," activity 2 as "enacting science" and activity 3 as "reflection."

Table 1. Differences between teachers' epistemologies about NOS

| Codes | Interviews | Enactment | Reflection |
|---|---|---|---|
| *A. Science* | | | |
| *i. The result of science:* | | | |
| 1. improves life quality | 14.3% | 0.4% | 0% |
| 2. is a number of laws that govern the physical world | 5.7% | 83.9% | 8.6% |
| 3. provides answers to questions about physical phenomena | 20.0% | 0% | 0.8% |
| 4. is an interpretation of everyday physical phenomena | 60.0% | 15.7% | 90.6% |
| *ii. Science (as a process) is:* | | | |
| 1. a process of memorizing scientific facts | 0% | 0% | 0% |
| 2. a process of observations and experimentation | 45.7% | 17.6% | 43.2% |
| 3. a process of investigating, accepting or rejecting theories or hypotheses | 54.3% | 82.4% | 56.8% |
| *B. Scientific knowledge is:* | | | |
| 1. absolutely true (there is only one answer) | 20.0% | 5.3% | 10.5% |
| 2. how each individual understands the truth about the physical world | 5.7% | 77.5% | 33.7% |
| 3. the most commonly accepted answer to questions about the physical world | 5.7% | 9.0% | 47.1% |
| 4. the knowledge we learn from experimentation | 54.3% | 4.1% | 6.2% |
| 5. a truth that might change in the future | 14.3% | 4.1% | 2.5% |

### Pre-service teachers' epistemologies related to what is science

When "enacting science" most of the pre-service teachers (83.9%) felt that the result of science consists of a number of laws that govern the physical world. This idea appeared less often in the interviews (5.7%) and in the reflections (8.6%). During the interviews, pre-service teachers suggested that science improves quality of life (14.3%) and provides answers to questions about physical phenomena (20.0%); these ideas did not appear during the two online discussions. During the interview and the reflection, the majority of the teachers felt that scientific knowledge is an interpretation of everyday physical phenomena (60% and 90.6%, respectively), an idea which did not appear as often when teachers enacted science (15.7%).

### Pre-service teachers' epistemologies related to what is the result of science

During the interviews, pre-service teachers described science as a process of observation and experimentation (45.7%). Similarly, this idea appeared in teachers' reflections (43.2%), while only 17.6% of the pre-service teachers seemed to use this idea while enacting science. When enacting science, most of the pre-service teachers (82.4%) seemed to view science as the process of investigating, accepting or rejecting theories. However, during their reflection and the interview, only about half of the pre-service teachers (56.8% and 54.3%, respectively) seemed to view science as such. The other half (43.2% and 45.7%, respectively) viewed science as a process of observation and experimentation. None of the teachers appeared to see science as a process of memorizing scientific facts during the three activities.

### Pre-service teachers' epistemologies related to what is scientific knowledge

While enacting science 77.5% of the pre-service teachers felt that scientific knowledge reflects how each individual understands the truth about the physical world, whereas, during the reflection fewer teachers (47.1%)

seemed to understand scientific knowledge as the most commonly accepted answers to questions about the physical world. During the interview more than half of the teachers (54.3%) indicated that scientific knowledge is what we learn from doing experiments. This idea appeared less during the rest of the activities. Moreover, during the interview 20,0% of the teachers seemed to understand scientific knowledge as absolutely true, but less seemed to do so during their reflections (10.5%) and even less when they enacted science (5.3%). Similarly, during their interviews 14.3% of the pre-service teachers described scientific knowledge as a "truth" that might change in the future, an idea observed much less during the enactment of science (4.1%) and the reflection (2.5%).

### Comparison of pre-service teachers epistemologies across the three contexts

Overall, the Wilcoxon Signed-Rank Test indicated that the pre-service teachers' epistemologies for each of the three aspects of NOS investigated were statistically significantly different across the three contexts. Specifically, the pre-service teachers' epistemologies about the result of science differed significantly among all three contexts ($p<0,001$ across all comparisons). In terms of what science is, pre-service teachers' epistemologies during the interview and the reflection were not found to differ significantly, but both were found to differ significantly from the enacting science activity ($p<0,001$ across all comparisons). Finally, the findings concerning what scientific knowledge is, revealed that all the contexts were significantly different ($p< 0,001$ across all comparisons).

## Discussion

The aim of this study was to investigate and describe pre-service teachers' epistemologies from three different data sources reflecting three different contexts. Findings revealed statistically significant differences of the pre-service teachers' epistemologies among the three contexts investigated, which raises two issues.

The first issue is related with the nature and/or the development of epistemological understanding. Our findings, along with a number of other studies (Hammer, 1994; Roth & Roychoudhury, 1994; Solomon, Duveen, & Scott, 1994; Hammer & Elby, 2002; Louca et al, 2004) contend that personal epistemologies seem to be better understood as fine-grain cognitive structures that are activated depending on the context in which the activation takes place. Asking teachers about their views of NOS or inferring those views from their enactment in science or their reflections seem to invoke statistically different epistemological resources, which they use to explain, understand or respond to a particular situation.

The fact that the differences among the three contexts are not similar in terms of the three areas of NOS that we have investigated (what is science, what is the result of science and what is scientific knowledge), has another important implication. It suggests that viewing personal epistemologies as semi-independent beliefs (Hofer & Pintrich, 1997; Schommer-Aikins, 2004) cannot explain these differences. The teachers' epistemologies investigated in this study, differ among the three contexts, but the differences in the three areas of NOS investigated are not similar among the three contexts. Teachers' epistemologies are not only different across contexts, but there is also a variation of those differences within the various contexts. For instance, pre-service teachers' views about science as a process are similar between the interview and the reflection. However, the same teachers' views about the nature of scientific knowledge are statistically significantly different in the two contexts. Thus, we suggest that our findings indicate that even within a particular area of NOS, the same pre-service teachers hold multiple epistemological resources, which they activate based on context. Thus, attributing robust epistemologies developed into coherent theories (McComas, 1996; Sandoval & Morrison, 2003) to pre-service elementary teachers is definitely inappropriate.

The second issue is related to the debate that exists in literature about the approaches of studying personal epistemologies (Sandoval, 2005). Assessments of formal science tend to ask students to express their views about the nature of scientific knowledge and activity, including questions about what scientists do, what theories are, how theories and experimentation influence each other and so forth. The general picture from such studies is that students' ideas about formal science follow a developmental trajectory toward increasing sophistication throughout adolescence (e.g., Leach et al., 1997), but tend to remain fairly naïve even during university instruction. Our findings show a similar picture, which we feel suggests that the research community is far from settling the debate as to which particular approaches should be used to assess or study personal epistemologies. Depending on the context and the manner of investigation, students and teachers may "show" different epistemological understanding. One suggestion is that, the research community may choose to follow a multiple-data-source approach when studying personal epistemologies.

Finally, disagreeing with the suggestion that it is an academic issue whether teachers' epistemologies are necessarily reflected in their classroom practice (Lederman, 1999), an implication stemming from our findings is that methodological effects might be possible. The ideas teachers express, explicitly or implicitly, may differ depending on the situation in which they are engaged. For example in the context of an interview, teachers may report views they believe they hold, whereas when they enact science, or reflect back on their enactment they may reveal "alternative" views. In the interview context, the teachers self-report views they think they hold about an issue, which are often influenced by social desirability. While enacting science their conceptions will be interpreted by independent coders and are possibly influenced, not only by what they have been taught about science, but also by their experience with science in formal informal learning contexts. Consequently, we propose that a broader account of personal epistemologies should include investigations with methods that are not limited to questionnaire surveys and interviewing.

## References


Abd-El-Khalick, F. & Boujaoude, S. (1997). An exploratory study of the knowledge base for science teaching. *Journal of Research in Science Teaching, 34*, 673-699.

Abd-El-Khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of the nature of science: A critical review of the literature. *Int. J. of Science Education, 22*(7), 665–701.

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education, 82*(4), 417–436.

Abell, S, Martini, M. & George, M. (2001). That's what scientists have to do: preservice elementary teachers' conceptions of the nature of science during a moon investigation. *International Journal of Science Education, 23*(11), 1095 - 1109

Abell, S. K. & Smith, D. C. (1994). What is science?: Preservice elementary teachers' conceptions of the nature of science. *International Journal of Science Education*, 16, 475- 487.

American Association for the Advancement of Science (1990) *Science for all Americans.* New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for scientific literacy.* New York: Oxford University Press.

Bell, P., & Linn, M. C. (2002). Beliefs about science: How does science instruction contribute. In B. K. Hofer & P. R. Pintrich (Eds.), Personal epistemology: The psychology of beliefs about knowledge and knowing (pp. 321-346). Mahwah, NJ: Lawrence Erlbaum.

Brickhouse, N. W. & Bodner, G. M. (1992). The beginning science teacher: classroom narratives of convictions and constraints. *Journal of Research in Science Teaching*, 29, 471-485.

Calderhead, J., & Robson, M. (1991). Images of teaching: Student teachers' early conceptions of classroom practice. *Teaching and Teacher Education, 7* (1), *1-8.*

diSessa, A. A., Elby, A., & Hammer, D. (2002). J's epistemological stance and strategies. In G. M. Sinatra & P. R. Pintrich (Eds.), Intentional conceptual change (pp. 237-290). Mahwah, NJ: Lawrence Erlbaum.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological Review, 87(6), 477-531.

Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction*, 12(2), 151–183.

Hammer, D. M., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 169–190). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hodson, D. (1993) Philosophic stance of secondary school science teachers, curriculum experiences, and children's understanding of science: some preliminary findings. *Interchange*, 24, 41-52.

Hofer, B. K. (2002a). Epistemological world views of teachers: From beliefs to practice. Issues in Education, 8(2), 167-173.

Hofer, B. K. (2002b). Personal epistemology as a psychological and educational construct: An introduction. In B. K. Hofer & P. R. Pintrich (Eds.), Personal epistemology: The psychology of beliefs about knowledge and knowing (pp. 3-14). Mahwah, NJ: Lawrence Erlbaum.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Ed. Research, 67,* 88–140.

Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching*, 39(7), 551– 578.

King, P. M., & Kitchener, K. S. (1994). Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults. San Francisco: Jossey-Bass.

King, P., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. Educational Psychologist, 39, 5–18.

Leach, J., Millar, R., Ryder, J., & Sere, M.-G. (2000). Epistemological understanding in science: The consistency of representations across contexts. *Learning & Instruction*, 10, 497– 527.

Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29, 331-360.

Lederman, N. G. (1999). Teachers' understanding of the nature of science and classroom practice: factors that facilitate or impede the relationship. *Journal of Research in Science Teaching*, 36, 916-929.

Linn, M. C., & Songer, N. B. (1993). How do students make sense of science? Merrill-Palmer Quarterly, 39(1), 47 – 73.

Luft, J. A., Roehrig, G. H., & Patterson Nancy C. (2003). Contrasting landscapes: A comparison of the impact of different induction programs on beginning secondary science teachers' practices, beliefs, and experiences. *Journal of Research in Science Teaching*, 40(1), 77-97.

Louca, T. L., Elby, A., Hammer, D. & Kagey, T. (2004). Epistemological resources: Applying a new epistemological framework to science instruction. *Educational Psychologist 39(1)*, 57-68.

McComas, W. F., & Olson, J. K. (1998). The nature of science in international science education standards documents. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 41–52). The Netherlands: Kluwer Academic Publisher.

Millar, R. & Osborne, J. (1998). *Beyond 2000: Science Education for the Future.* London: King's College.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academic Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K*-8. Washington, DC: Committee on Science Learning, Kindergarten Through Eighth Grade.

Palmquist, B.C., & Finley, F.N. (1997). Preservice teachers' views of the nature of science during a postbaccalaureate science teaching program. *J. of Research in Science Teaching*, 34, 595–615.

Penso, S., Shoman, E., & Shiloah, N. (2001). First Steps in Novice Teachers' Activity. *Teacher Development, 5*(3), 323 - 338.

Piaget, J., & Foundation, N. F. (1967). Further aspects of Piaget's work. London: Routledge & K. Paul.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), Handbook of research on teacher education (pp. 102 ± 119), New York: Macmillan.

Roth, W. M. & Roychoudhury, A. (1994). Physics students' epistemologies and views about knowing and learning. *Journal of Research in Science Teaching*, 31(1), 5 –30.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Wiley Periodicals, Inc.*

Sandoval, W. A., & Morrison, K. (2003). High school students' ideas about theories and theory change after a biological inquiry unit. *Journal of Research in Science Teaching*, 40(4), 369 – 392.

Schommer-Aikins, M. (2004). Explaining the epistemological belief system: Introducing the embedded systemic model and coordinated research approach. *Educational Psychologist, 39,* 19–29.

Solomon, J., Duveen, J., & Scott, L. (1994). Pupils' images of scientific epistemology. *International Journal of Science Education*, 16(3), 361– 373.

Tobin, K., & McRobbie, C. (1997). Beliefs about the nature of science and the enacted science curriculum. Science and Education, 6(4), 355-371.

Tobin, K., Tippin, D. J. & Gallard, A. J. (1994). Research on instructional strategies for teaching science. In D. L. Gabel (Ed.), Handbook of research on science teaching and learning. New York: National Science Teachers Association.

Van Driel, J., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. *Journal of Research in Science Teaching*, 38, 137-158.

# Exploring how novice teachers learn to attend to students' thinking in analyzing case studies of classroom teaching and learning

Daniel M. Levin, American University, 4400 Mass Ave., NW, Washington, DC 20016, levin@american.edu
Jennifer Richards, University of Maryland, College Park, College Park, MD 20742, jrich@umd.edu

**Abstract:** In this paper, we explore how candidates in a science teacher preparation cohort attend to the substance of student thinking while watching classroom videos or reviewing students' written work. Our findings suggest that the teacher candidates are able to attend to specific student ideas and reasoning from the beginning of their pre-service preparation, but their practices of attending become more sophisticated over time. We also consider participation dynamics within the cohort, as participants assume different roles and begin to regulate their discussions.

## Introduction

The National Research Council (NRC) (2007) characterizes students' science learning in terms of knowledge and use of conceptual content, reasoning abilities, epistemological understandings, and participation in scientific practices. In line with the NRC's conceptualization of science education reform, it is suggested that "proximal formative assessment" (Erickson, 2007), as it refers to teachers' ongoing, everyday attention to the substance of students' ideas, plays an important role in shaping teachers' instructional moves and supporting students' science learning across these strands (Atkin & Coffey, 2003). We believe that an important focus for science teacher education is thus to help novice science teachers learn to attend to the substance of student thinking, interpreting the meaning students are trying to convey.

This paper is our effort to document what happens in our science pedagogy program, which takes as a specific aim the development of teacher candidates' practices of attending to the substance of student thinking while examining records of classroom practice. We report results from the first two semesters in the science pedagogy course sequence, focusing our inquiry on three questions:

- What do our teacher candidates attend to when discussing records of classroom practice?
- How do our candidates attend to student thinking?
- How do practices of attending to student thinking develop over time within the cohort?

## What teachers notice in records of classroom practice

We refer to practices of "attending" to student thinking, but our work is similar to a body of literature primarily in mathematics education that uses the term "noticing." The noticing literature is explicitly focused on the substance of student thinking, responding to reform documents in both mathematics and science education (NCTM, 2000; NRC, 2007) that call for teachers to "base their instruction on the lesson as it unfolds in the classroom, paying particular attention to the ideas that their students raise" (van Es & Sherin, 2008, p. 244). Several scholars (Hammer & Schifter, 2001; Jacobs, et al., 2007) argue that professional development and teacher education aimed at focusing teachers' attention on the substance of student thinking is crucial for teacher learning; it is assumed that helping teachers notice students' ideas when exploring records of practice like classroom videos (e.g., van Es & Sherin, 2008) and samples of student work (e.g. Kazemi & Franke, 2004) will amplify teachers' tendencies to do so in their own classrooms. For the purposes of this paper, we are focusing on the teacher education setting, but we take up the issue of connections between teacher education and teachers' classroom practices in our discussion.

Little research has been conducted on pre-service teachers' practices of attending to student thinking, but Carter et al. (1988) suggest that novice teachers' abilities to notice student thinking are poorer than experienced teachers' abilities. Theoretically, lacking any experience in classrooms, new teachers have more difficulty hearing and interpreting student ideas in the classroom than experienced teachers do. However, in a more recent study in a pre-service secondary mathematics teacher education course, Star and Strickland (2008) found that teacher candidates generally did not enter the course with well-developed observation skills, but the course led to significant increases in these skills, particularly novice teachers' abilities to notice features of the classroom, mathematical content, and student thinking. Sherin and van Es (2005) have also shown that pre-service teachers can learn to attend to student thinking fairly quickly. Our current study contributes to the noticing literature and literature on pre-service teacher education by exploring what happens in a pre-service science teacher pedagogy course sequence focused on attending to the substance of student thinking.

## A framework for learning to attend to the substance of student thinking

Drawing on the noticing literature, and based on our iterative coding for this project (see the section on data analysis below), we describe in this section what we take as evidence of attending to the substance of student thinking. We also draw from other genres of literature, including physics education (diSessa, 1993; Hammer, Elby, Scherr, & Redish, 2005), cognitive science (Minsky, 1985), sociolinguistics (Goffman, 1974; Tannen, 1993), and anthropology (Lave & Wenger, 1991) to understand what it means to learn to attend to the substance of student thinking.

### Attending to the substance of student thinking

Three aspects of attending to student thinking are prevalent in the noticing literature; we have identified these aspects in this study as well, and we consider them to provide varying degrees of evidence of attending to student thinking. These three aspects include identifying students' ideas and reasoning, interpreting the meaning students are trying to convey, and evaluating the ideas and reasoning inferred from students.

Goodwin (1994) describes "highlighting" or identifying ideas as an important part of what practitioners in a field do. Identifying important ideas helps to "divide a domain of scrutiny in a figure and ground, so that events relevant to the activity of the moment stand out" (Goodwin, p. 610). We consider identifying important ideas to be a necessary precursor to attending to the substance of those ideas.

Once teachers identify important student ideas, Crespo (2000) distinguishes between teachers' comments that are evaluative and focused on correctness and those that are interpretive and focused on understanding. We believe that van Es and Sherin's (2008) definition of interpreting is closest to the meaning that we ascribe to the term. As van Es and Sherin state, "… we want to emphasize the importance of interpreting classroom events. Thus, how individuals reason about what they notice is as important as the particular events they notice" (p. 247). We speak of attending to the substance of student thinking in this strict sense – interpreting the meaning students are trying to convey, without simply evaluating the ideas. Thus, we take interpretive statements to be the best evidence that a teacher is attending to the substance of student thinking. Furthermore, we believe that interpretive statements are the most productive in professional development contexts or teacher education courses – when teachers identify and interpret specific student ideas in collaboration with others, they have the opportunity to argue about their interpretations of the ideas, which leads to better-warranted evaluations and proposed instructional responses.

It is important to note that we view these aspects of attending to student thinking as analytical tools that help us make sense of how candidates are attending to students' ideas and reasoning. We are not making claims that these are separate cognitive processes within teachers' minds. Our purpose in describing these components is simply to examine those aspects of attending to student thinking that the candidates make explicit.

### Learning to attend to the substance of student thinking

Our perspective for understanding how teachers learn to attend to the substance of student thinking draws from research on learning in physics. Hammer (2000) argues that students do not draw on nascent fully formed theories to reason in physics but rather that students employ small-grained, context-sensitive *resources* to do so. Hammer's framework builds on diSessa's (1993) description of phenomenological primitives or "p-prims," which are conceptual resources, based on learners' experiences with physical phenomena, which can be useful for learning physics. Hammer et al. (2005) have expanded the idea of resources to include fine-grained bits of declarative and procedural knowledge, metacognition, epistemology, and understandings of social norms that are derived from people's past experiences and activated in different situations.

Hammer et al. (2005) also suggest that in any moment, locally coherent sets of resources or *framings* are activated that are mutually consistent and reinforcing. Framing stems from a diverse history in cognitive science and sociolinguistics (Goffman, 1974; Minsky, 1985; Tannen, 1993). Here, we define framing as an individual or collective sense of "What is going on here?" Thus, framing involves an interaction between the contextual cues present in any given situation and the resources that various participants already have.

In any practice, which Wenger (1998) refers to as sustained engagement in a joint enterprise with shared tools, newcomers must learn relevant norms. Lave and Wenger (1991) describe the process by which newcomers learn these norms as "legitimate peripheral participants" (p. 29), which we take to mean that newcomers learn the framing of a particular practice through engaging and participating in that practice. We will argue that our teacher candidates draw on resources that they already have in order to participate in the practice of attending to student thinking. Additionally, by engaging and participating in this practice in various ways, the cohort establishes and reinforces a collective framing of attending to student thinking.

The language of framing has been used to understand how physics students frame what is going on in a particular context and how their framing is associated with their physics learning (Hammer et al., 2005); it has only been recently applied to secondary science pre-service teacher learning. Additionally, there has been little focus on how a framing of attending to student thinking is established and reinforced in a community of preservice teachers. We return to our framework in our conclusion to articulate how pre-service secondary science teaching candidates learn to attend to student thinking while watching video and looking at samples of student work.

## Research context and methods

Our data come from the first two courses of the three-course science pedagogy sequence in a one-year graduate-level initial teaching certification program at the University of Maryland, College Park. The course sequence is explicitly structured to draw teacher candidates' attention to the substance of student thinking, first by having them collectively examine records of classroom practice (videos and samples of student work), and then by having them collect and analyze such records from their own classrooms. During the first course (summer session), the teacher candidates identify frameworks for understanding students' science learning in the literature, interview students about science topics, engage in their own scientific inquiry, examine curricula for opportunities to draw out students' ideas and reasoning, and discuss samples of student thinking in classroom video and student work. The second pedagogy course (fall semester) continues these practices of examining and discussing samples of student thinking but goes beyond the first course in helping candidates develop instructional strategies consistent with science education reform and respond to student ideas as they arise during instruction. Candidates write lesson plans in which they anticipate what students might say or do and how they (as teachers) might respond instructionally. They then teach these lessons, collect student work or recordings of the class, and analyze the student thinking in evidence. The third course involves candidates in collecting data from their own teaching and writing case studies of the student thinking in evidence. Here, we report on data from the first two pedagogy courses – specifically on how the teacher candidates attended to the substance of student thinking in records of classroom practice.

### Subjects

The first course was comprised of eleven pre-service secondary science teacher candidates. Seven of these candidates were in a one-year program to earn a Masters degree and certification. Three were post-doctoral scientists pursuing certification only, and one was a former patent attorney who was pursuing certification only. In the second course, three additional candidates joined who were participants in an integrated bachelors/Masters program for certification. They had taken the initial pedagogy course the previous year as undergraduates. For the purposes of this paper, we primarily consider the candidates as a group. We discuss some differences among candidates in terms of their participation, but an in depth discussion of other differences is beyond the scope of this paper.

### Procedures

We shared eight cases of secondary science classroom work with the teacher candidates. Six cases were videos (20-45 minutes long) of secondary science classrooms with typed transcripts and/or captions. One of the videos was shown twice, as we discuss below. Two cases were collections of samples of student work. We selected all of the cases from a collection developed as part of another project (Levin, 2008); these cases will be included in a book/DVD package of teacher-authored case studies, similar to one produced for elementary teacher education (Hammer & van Zee, 2006).

As the instructor, Levin began the discussion of each case by describing the context in which the work occurred or by having the group read the introduction to the teacher's written case study. We then shared the video or student work with the group, and Levin asked, "What do you notice in the students' ideas and reasoning?" Levin facilitated the discussion to draw specific attention to the substance of students' ideas and reasoning. For example, if candidates made a general statement such as, "It seems like the students get it," Levin would say, "Can you point to something someone said or did that makes you think they get it?" Similarly, if candidates directed attention to the action of the teacher by suggesting what the teacher should do or describing problems with the teacher's approach, Levin would ask what they saw in the students' reasoning that led them to make that claim about the teacher.

### Data collection

We videotaped the candidates' discussions of student thinking in each of the cases. The discussions were each approximately 30-45 minutes in length. Due to the great variability among the cases, it was difficult to compare the cases in order to explore how the cohort's practices of attending to the substance of student thinking changed over time by looking at the progression throughout the cases. Thus, we showed the case we had shown at the beginning of the summer session again at the end of the fall semester in order to look at differences in how the candidates attended to the substance of student thinking at the beginning of the first course and at the end of the second.

### Data analysis

To explore our first question ("What do our teacher candidates attend to when discussing records of classroom practice?"), we drew on a coding scheme to categorize each speech turn, which was developed by inductive coding (Miles & Huberman, 1994) in a similar project with practicing teachers (Levin, 2008). We then developed our codes further through an iterative process of coding a sample of the discussions, discussing our codes, and expanding or collapsing codes as appropriate. In doing so, we developed a scheme that organized what candidates attended to into eight categories: specific student thinking, general student thinking, the actions of the teacher, the nature of the activity, the science content, student attributes, student engagement, and "other." We provide a detailed description of how we defined each of the codes in another paper (Levin & Richards, 2009), but the most important categories for our purposes in this paper are specific student thinking and general student thinking. We identify candidates' comments about specific student ideas or reasoning as attending to specific student thinking; for instance, we would code a statement like, "Maybe he's saying that they are practice for hunting – the snakes are just practice" as attending to specific student thinking. Alternately, we identify candidates' comments about the general understanding or reasoning of students in the class as attending to general student thinking (e.g. "I think most of them get it"). Levin coded all of the transcripts. Richards coded one third of the transcripts, and we compared our coding to arrive at an inter-rater reliability of 83%. We then discussed each disagreement until we reached consensus on the remaining codes.

To explore our second question ("How do our candidates attend to student thinking?"), we conducted another round of iterative coding, focusing only on the comments we had coded as attending to specific student thinking. As we previously described, we saw that candidates made three kinds of comments that we coded as attending to specific student thinking. At times, candidates 1) simply identified students' ideas, which we took as evidence that the candidates noticed the ideas, but we could not tell whether they attended to the meaning that the students were trying to convey. Candidates also 2) made evaluative statements in reference to students' ideas, which again indicated that the candidates were attending to the ideas that were present and perhaps making tacit interpretations of these ideas, but their interpretations (if present) were not made public. Finally, candidates sometimes 3) attempted to interpret what students were saying, which we took as the strongest evidence that they were attending to the substance of students' ideas. We did not take frequency counts of this coding because many utterances included overlapping codes at this level of analysis; instead, we considered how the three kinds of comments seemed to be related in candidates' statements. We explore the occurrences and relationships of these codes qualitatively in our data analysis.

To explore our third question ("How do practices of attending to student thinking develop over time within the cohort?"), we showed candidates the same case twice – once at the beginning of the summer session and again at the end of the fall semester. This case, hereafter referred to as the "Owls and Snakes," showed a teacher and students discussing a strange relationship between a species of owl and a species of blind snake that lives in the owls' nests undisturbed. We compared the initial coding of the first showing with the coding of the second showing. The discussions were slightly different in length, so we normalized the results by converting them to the frequency of codes uttered per 30 minutes of discussion. Both of us scored both transcripts completely, and we had 84% inter-rater agreement. Again, we resolved differences in coding by meeting and agreeing on the disputed codes.

We also looked across all of the case study discussions for patterns in the nature of the conversations and how participation in the norms and practices of attending to student thinking developed over time within the cohort. Specifically, we looked at who participated and how they did so over the course of the two semesters, how candidates drew each other's attention to specific student ideas, and how the role of the facilitator changed.

## Data and analysis

In this section, we present the results of our coding, followed by our analysis of the data and a discussion of the cohort's practices over time.

### Results and analysis of coding

We collected 995 coded passages over the course of nine discussions, including discussions of the eight cases plus the repeat of the first case. We coded 43% of utterances as specific attention to student thinking, 6% as general attention to student thinking, 18% as attention to teacher action, 9% as attention to the activity, 11% as attention to science content, 3% as attention to student attributes, 1% as attention to student engagement, and 6% as "other."

As we discussed above, it was difficult to make any claims about changes in what candidates attended to because of differences in the content of the cases. To look at changes, we led a discussion of the same case study at the end of the second course that we had discussed at the beginning of the first course and compared the changes in the frequencies of our coding. We only saw notable changes in the relative frequencies of attending to "specific" and "general" student thinking – the percentage of specific comments about student thinking (per

30 minutes) increased from 36% to 48%, while the percentage of general comments about student thinking decreased from 7% to 2%.

Thus, we can make two assertions related to our research questions at this point. First, our secondary science teacher candidates were able to attend to the substance of student thinking from the beginning of the pedagogy course sequence. Second, our candidates focused more on specific student ideas relative to general student ideas over time.

We saw that candidates consistently attended to the substance of student thinking from the beginning of the first pedagogy course. We coded almost half of all comments throughout the cases as "attending to specific student thinking," and our coding varied little across the cases in this respect. In terms of how candidates attended to student thinking, we saw that candidates routinely identified, interpreted, and evaluated students' ideas. At times, candidates made comments simply identifying a student's idea, and Levin followed up to ask what the candidates thought the student meant and what the candidates thought of the idea. Frequently, however, candidates specifically interpreted the student's meaning without prompting. These specific interpretations frequently led to sophisticated evaluations of students' conceptual understanding, reasoning, epistemological stances, and participation in scientific practices. Specific interpretations often occurred during long stretches of conversation that were about students' ideas and reasoning. For example, during an early case discussion in which candidates were discussing whether the students understood the relationships among force, mass, and acceleration when considering gravitational motion, Sarah, who was often one of the quieter students, identified an idea on a student's worksheet that she did not understand:

> Sarah: "I was confused by what she meant about inertia canceling out, like for, on page 2, when they talked about how... and not falling at the same time because their inertia's different?"
>
> Jack: "Well again I think that just mass, or heavier mass is less acceleration because they were just going back to that and less mass is higher acceleration."
>
> Alex: "It's interesting because on question 3 she -- at first the student states the right answer, they've got the concept that they land at the same time, and she understands that things that fall land at the same time, but then has trouble explaining why... she has this idea of the inertias canceling each other out, which indicates that she doesn't really have an understanding of what inertia is or how it applies in the case of falling objects."
>
> Jack: "Well again I think that goes back to their thinking the forces are the same, because she's saying 'the higher the mass the lower acceleration' versus 'a lower mass and a faster acceleration,' they are going to equal the same thing, so that's what she means, 'canceling out' -- they're gonna equal the same thing."

Here, Alex offered an interpretation of the student's thinking that she "has this idea of the inertias canceling each other out" and evaluated that she "doesn't really have an understanding of what inertia is or how it applies in the case of falling objects." Jack interpreted the student's idea more specifically, suggesting why the student might be thinking about "canceling out," which had not been obvious to everyone. As the conversation continued, Mark suggested another possible interpretation for what the student was thinking, and Elsa, Sarah, Ryan, and Alex debated Mark's interpretation, all drawing on the student's responses to other problems to debate what she might mean by "canceling out." This conversation about one specific student's thinking was followed by a discussion about how to teach the $F = ma$ formula more generally, including how to help students recognize different situations (e.g. when acceleration is constant versus when force is constant).

However, candidates often made general comments about student thinking early in the first semester. In the first discussion of the "Owls and Snakes" case, for example, we heard many general claims about what students understood or how they were reasoning. Candidates made declarative statements like "They were thinking out loud, and thinking logically;" "They're asking the right questions;" and "They're doing good stuff, they're reasoning, they're connecting their prior knowledge" without including interpretations of students' specific ideas to warrant their statements.

Candidates also made general comments about student thinking after long discussions of particular students' ideas. For example, in the "Owls and Snakes" case, the teacher presents the students with some data, which leads to an interesting argument about whether or not the data fits students' hypothesized relationship between the owls and snakes. During the first viewing of the case, candidates had a long discussion about particular students' ideas during this segment (including a number of comments we coded as "specific" student thinking), at the end of which Ryan made the following general claim:

> "I can see the students are, uh, doing something that I agree with, which is not assuming that just because there's data that indicates something, that that means that [the owls and snakes are] getting something out of it."

Here, we see an example of a general statement about student thinking, which is an evaluation that the students are doing something with which Ryan agrees. It differs from the other general statements above in that, following a conversation about specific student ideas and reasoning, it is grounded in the interpretations that candidates provided during the preceding conversation. Ryan was therefore able to specify what he liked about the students' argument – they did not assume that the data supported a particular answer.

When we looked at candidates' general comments in the second discussion of the "Owls and Snakes" case, we found that there were fewer general comments relative to specific comments. Also, the general comments were all of the kind that followed interpretations of specific student ideas and reasoning; none were the blanket evaluations we had seen in the first discussion. This kind of general claim about student thinking draws on specific interpretations of students' ideas and reasoning and can therefore provide novel and productive warrants for the evaluation of student thinking.

## Developing practices of attending to student thinking

In addition to exploring the content of what candidates attended to, we looked at how their participation in the conversations changed over time. In the earliest discussions, four candidates spoke the most in the discussions: Alex (the patent attorney), Brian and Elsa (post-doctoral scientists), and Ryan (an engineer by training). Alex in particular seemed to understand that the central aspect of the practice of attending to student thinking was to make claims about students' meaning by identifying specific things that students said. As we discussed above, Levin actively modeled this practice by asking candidates for specific examples. In the exchange below from the first "Owls and Snakes" discussion, we see Alex jumping in with an example even before Levin has finished asking for it:

> Ryan: "I thought it was a really impressive class."
>
> Levin (instructor): "Say more about that, why?"
>
> Ryan: "Because, uh, they were thinking out loud, and thinking logically, and the teacher was doing a great job of getting them to use reasoning."
>
> Levin: "So let's see if we can find – " (overlapping with Alex)
>
> Alex: (overlapping) "I like that distinction of that there's the good, the good maid and the bad maid, because the students are told there's a distinction, right, some snakes are eaten and some snakes aren't, there's eighty-nine percent that are alive and eleven percent that are dead, although they're not really told that they're eaten. Only one seems to be half eaten. So they're trying to immediately come up with a reason about why there are these two groups, why some snakes are alive and some snakes are dead and the reason that they come up with, well some are good at burrowing and cleaning up the nest and some are bad at this job and so they're eaten by the owls. At least that's an interesting reaction to being told there's two groups and they immediately come up with some mechanism, some reason, some logical reason to explain why there are two groups, why there are alive and dead snakes."

Throughout the case study discussions, Alex continued to identify examples of student thinking himself and to provide interpretations of others' examples to support evaluations. In some cases, he asked other candidates to support their statements with references to the transcripts or student work, asking several times "Where is that?" or "Where do you see that?" Alex appeared particularly comfortable with practices of interpreting students' meaning. Both in and outside of the pedagogy classes, Alex explained that his work as a lawyer helped him to focus on what people were saying.

Although Alex and some others dominated the conversations at the beginning, others began to participate within the first few case study discussions. By the second viewing of the "Owls and Snakes" case in the second pedagogy course, multiple candidates were participating in long conversations about student thinking without prompting from Levin. For example, Jack brought attention to a situation in the video in which a student, responding to a question from the teacher, said that a particular piece of data could be "used to evaluate" the students' hypotheses about the relationship between the owls and snakes. Jack thought that the student was just choosing one of the options the teacher had given him (can or cannot be used to evaluate the hypotheses) without thinking about it. Other candidates were not so sure. When the teacher asked the student, "What does it add that will help us answer the question?" the student replied:

> *"Uh, it could enter the nest on its own. The snakes are capable of climb, climbing up trees, and they can get to the nests on their own… and, if owls and snakes ever turn against each other, they could use that as an advantage for like, uhhh, battle and stuff."*
>
> Kay: "He's saying that the snakes are making the choice to go there. Like I…"
>
> Mark: "Right, that's what I read here."
>
> Kay: "Right, and then that's what's important about it to him…"
>
> Jack: "Right, so right. I see where that is important but I don't know if [he] understands that… he says, 'Yeah, they can climb trees'… but then line 23 I don't understand what he was meaning there because [he's] like, 'If they ever turn against each other they can use that to their advantage.'"
>
> Maria: "He probably means they could just climb back down to escape."

Kay interpreted the importance the student was placing on choice, but Jack questioned whether the student understood why it was an important idea. Maria interpreted the student's idea about why climbing trees was advantageous to the snake. The point of this snippet is not to definitively interpret and evaluate the student's idea. The point is that the candidates were focused on trying to understand the student's unconventional idea and truly attending to his meaning -- with little scaffolding from Levin nor the participation of Alex and other participants who had dominated the earlier conversations about student thinking.

## Conclusions and future study

This work contributes to a growing understanding of novice teachers' abilities to attend to the substance of student thinking and provides insight into how their practices of doing so develop and change over time. We show evidence that novice teachers can identify, interpret, and evaluate the substance of student thinking when they participate in pedagogy courses designed to draw attention to this topic. Additionally, the candidates in our pedagogy courses became more adept at drawing specific interpretations of student thinking and using these interpretations to warrant general claims and evaluations of student understanding.

However, a particularly important aspect of our data is that the cohort of teacher candidates attended to student thinking from the very beginning of the pedagogy course sequence. We believe that this finding has important implications in terms of how to best conceptualize learning to attend to the substance of student thinking. The predominant assumption in the noticing literature is that noticing is a "skill" (e.g. Jacobs et al., 2007; van Es & Sherin, 2008). We challenge the use of the term "skill," which typically implies something that teachers do not know how to do until they have been taught. Even if the use of the term "skill" in the noticing literature is not intended in this manner, we argue that conversations about learning should always be explicitly connected to a strong theoretical base, and we do not believe that the term "skill" accurately captures what is learned or how learning occurs.

Our data from this study support the presence of resources that teachers have to attend to the substance of student thinking. To us, the existence of these resources is important because it suggests that teachers do not learn a new skill in professional development or teacher education contexts, but that these contexts instead activate resources that teachers already have. That is, focusing on the substance of student thinking is not something developed de novo, but rather an activation of the resources that all people have for listening to the meaning that another person is trying to convey. Thus, an important task for teacher education is to help pre-service candidates draw upon these resources to support a framing of teaching in terms of attending to student thinking.

We suggest that the framing of teaching in terms of attending to student thinking was not simply put in place by Levin, the instructor, but was supported collectively through interaction among the participants in the group. Alex and some other candidates, perhaps because of prior experiences in similar settings, entered into the conversations very quickly and helped to support the framing that Levin was trying to establish. Over time, Levin's voice became less prominent as candidates pushed each other to articulate the specific evidence in student thinking that warranted claims of students' reasoning and understanding. The spirit of these exchanges over case studies continued into the third semester of the program, where candidates presented case studies from their own classrooms.

Our findings also suggest productive avenues for further research. As we noted, there were differences in the ways that individual candidates attended to student thinking. Some, like Alex, offered substantive interpretations of students' ideas from the very beginning. Others had difficulty doing this at first and seemed to become better at it as they engaged more with the group and saw the practices modeled by other candidates. We are pursuing more in-depth case studies of particular candidates in order to address these distinctions.

We are also examining candidates' practices of attending to the substance of student thinking while they are teaching in their own classrooms, where they must listen to student ideas in real time while trying to manage other facets of the classroom and the curriculum. We have followed several candidates into the classroom, and we plan to continue following them through induction and the early years of their teaching careers in order to better understand how and when teachers attend and respond to the substance of student thinking while teaching science.

## References


Atkin, J. and Coffey, J. E., Eds. (2003). *Everyday assessment in the science classroom.* Arlington, VA, NSTA Press.

Carter, K., Cushing, K., Sabers, D., Stein, P., and Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education 39*(3): 25-31.

Crespo S. (2000). Seeing more than right and wrong answers: Propspective teachers' interpretations of students' mathematical work. *Journal of Mathematics Teacher Education 3:* 155-181

diSessa, A. A. (1993). Toward an epistemology of phyiscs. *Cognition and Instruction 10*(2/3): 105-225.

Erickson, F. (2007). Some thoughts on "proximal" formative assessment of student learning. *Yearbook of the National Society for the Study of Education*. G. D. Fenstermacher, National Society for the Study of Education. 186-216.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Cambridge, MA: Harvard University Press.

Goodwin, C. (1994). Professional vision. *American Anthropologist 96*(3): 606-633.

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics, Physics Education Research Supplement, 68* (1), 52-59

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. Maestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89-120). Greenwich, CT: Information Age Publishing.

Hammer, D. and Schifter, D. (2001). Practices of inquiry in teaching and research. *Cognition and Instruction 19*(4): 441-478.

Hammer, D. and van Zee, E. (2006). *Seeing the science in children's thinking: Case studies of student inquiry in physical science*. Portsmouth, NH, Heinemann.

Jacobs, V. R., Clement, L. L., Philipp, R. A., Schappelle, B., & Burke, A. (2007). Professional noticing by elementary school teachers of mathematics, *American Educational Research Association*. Chicago, Illinois.

Kazemi, E. and Franke, M. L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education* 7: 203-235.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: University of Cambridge Press.

Levin, D. M. (2008). *What secondary science teachers pay attention to in the classroom: Situating teaching in institutional and social systems.* Unpublished doctoral dissertation, University of Maryland at College Park.

Levin, D. M., & Richards, J. (2009). Developing a professional vision for science education reform teaching, *Paper presented at Annual Meeting of the Association for Science Teacher Education*. Hartford, CT.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks: Sage Publications.

Minsky, M. L. (1986). *Society of mind.* New York: Simon & Schuster.

NCTM (2000). *Principles and standards for school mathematics*. Reston, VA.

NRC (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC, Committee on Science Learning, Kindergarten through Eighth Grade.

Sherin, M. G. and van Es, E. A. (2005). Using video to support teachers' ability to interpret classroom interactions. *American Educational Research Association.* Montreal, Canada.

Star, J. R. and Stickland, S. K. (2008). Learing to observe: Using video to improve mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education 11*: 107-125.

Tannen, D. (1993). *Framing in discourse.* New York, Oxford University Press.

van Es, E. A. and Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education 24*: 244-276.

# Collaborative Productivity as Self-Sustaining Processes in a Grade 4 Knowledge Building Community

Jianwei Zhang, State University of New York at Albany, USA. jzhang1@albany.edu
Richard Messina, Institute of Child Study Laboratory School, Toronto, Canada. rmessina@oise.utoronto.ca

**Abstract**: This study elaborates collaborative productivity as self-sustaining processes along with the role of the teacher in a knowledge building community. The participants were 22 fourth-graders, who investigated light over a three-month period facilitated by a veteran teacher with the support of Knowledge Forum®. Content analysis of the students' portfolios indicates significant advancement of understanding. Qualitative analysis of classroom videos, online discourse, and the teacher's reflection journal identifies community interactions and teacher scaffolding related to four interrelated processes. These include: (a) accumulating a highly variable stock of information and ideas and mobilizing information connection; (b) sustained, incremental idea development; (c) critical examination and selection of ideas; and (d) distributed emergent control. These processes elaborate self-organization mechanisms underpinning collaborative productivity, informing new ways to scaffold knowledge building.

## Introduction

With knowledge-based organizations pervading a knowledge economy and society, education is facing an unprecedented demand of preparing students for collaborative and creative knowledge practices. Various inquiry-based collaborative learning programs have emerged to achieve high-level collaborative productivity in a learning community. Students engage in productive sharing, conversation, and collaborative problem solving that lead to deep understanding and advancement of collective knowledge (Barron & Darling-Hammond, 2008; Hmelo-Silver & Barrows, 2008). Efforts to explain and support collaborative productivity often focus on the teacher/designer's role in charting, organizing, and guiding the processes, with student collaborating in fixed groups following specific scripts. This focus on "strong leader" is further heightened by the argument that "minimally guided teaching techniques" do not work (Sweller et al., 2007). Alternative to this "strong leader" explanation is a self-organization perspective: collaborative productivity emerges from complex, distributed interactions with the teacher as an important participant; small intellectual input from the members builds on one another to enable increasingly complex work and emergent progress. This self-organization perspective gains support from the recent studies on creative research teams and professional communities (Dunbar, 1997; Sawyer, 2007). In line with these studies, our recent design research traced the improvement of knowledge building in a Grade 4 classroom. Over three years, the designs evolved from fixed-group collaboration that involved extensive teacher-coordination to distributed, opportunistic collaboration. The third year design led to most productive knowledge advancement (Zhang et al., 2009). The present study analyzes an extended iteration of the above design experiment. The goal is to provide an elaborated account of collaborative productivity as self-sustaining processes, which will inform new and expanded ways to scaffold knowledge building.

Our exploration of collaborative productivity as self-sustaining processes was inspired by recent research on creative communities, such as productive research labs (Dunbar, 1997) and innovative professional communities (Engestrom, 2008). A creative community works as a system that is further embedded in the larger systems in a field (Csikszentmihalyi, 1999). Collaborative knowledge creation in these communities unfolds as self-organizing and emergent processes. "[T]he most innovative teams are those that can restructure themselves in response to unexpected shifts in the environment; they don't need a strong leader to tell them what to do." (Sawyer, 2007, p. 17) The mechanism of self-organization "is basically the combination of an evolutionary raw material—a highly variable stock of information from which to select possible patterns—and a means for experimentation, for selecting and testing new patterns." (Meadows, 1999, p.15) In light of the notion of self-organization, we synthesize four essential processes that sustain knowledge creation drawing on the literature on creativity and knowledge creation.

(a) Accumulating a highly variable stock of information and ideas, with dynamic information flow. Creativity emerges when individuals, often working in teams, produce novel variations to the domain that are recognized by the field composed of peer workers (Csikszentmihalyi, 1999). Each domain (e.g., physics) evolves a public knowledge base representing the state-of-the-art understanding (Bereiter, 2002). The public knowledge space gives knowledge and ideas an out-in-the-world existence (e.g., in books). The current knowledge base is the evolutionary outcome of the past work and meanwhile thinking resources and devices for future advancement, with new advances rooted in the past and further informing ever deeper inquiries.

(b) Positive loops fueling idea generation and development: Positive feedback loops are self-reinforcing, with the output of one operation becoming the input to another. As Dunbar (1997) observed in high-performing research labs, researchers perform cognitive operations and pass the results on to peers, who then use the results as the input to further cognitive operations to create new theories and experiments. A series of small operations may lead to major, often unexpected advances, which often cannot be attributed to any individual. As a part of the positive loops, the knowledge gained about a topic helps knowledge workers to further identify what is not known and ask deeper questions (Miyake & Norman, 1979), which lead to further actions to advance knowledge (Bereiter & Scardamalia, 1993).

(c) Negative loops that enable critical examination and selection of ideas. Positive loops in a creative system need to be coupled with, and slowed down by, self-correcting negative feedback loops in order to keep important system states within safe bounds (Meadows, 1999). In academia, the negative loop is partly played out through blinded peer review and critical academic discourse. The specific criteria used to judge whether a contribution is an improvement/advancement differ across domains. Such criteria and related rules of knowledge work are further internalized by individuals and become a part of their reflective thinking and discourse, so that they can make smart decisions to tease out good ideas from bad ones and choose the most promising ideas to work on in a way that can be accepted by peers (Csikszentmihalyi, 1999; Sternberg, 2003).

(d) Distributed emergent control. Innovative teams perform spontaneously (Sawyer, 2007). Team members take on high-level collective cognitive responsibility: They collaboratively monitor their progress and emergent goals, develop and refine work processes and procedures, and group and re-group in the service of the emergent needs (Chatzkel, 2003). They invent and adopt various knowledge tools and artifacts to support their productivity and expand the scope and sophistication of thinking (Hakkarainen, 2009). They critically reflect on the social norms that specify accepted ways of conducting research and other knowledge practices, adjust existing norms, and strive for new paradigms as needed (c.f. Sternberg, 2003).

Using the self-organization mechanisms as a lens, this study views into a knowledge building community to understand how collaborative productivity occurs and can be enabled. In a knowledge building community, students engage in continuous idea improvement to advance the state of the art of the community's knowledge, mirroring the socio-cognitive dynamics of knowledge creation in the real world (Bereiter, 2002; Scardamalia, 2002). This process is augmented through technology-based environments such as Knowledge Forum, which provides a communal knowledge space and related interaction tools for knowledge building discourse (see Scardamalia, 2004). This study analyzes the knowledge building work of a Grade 4 classroom facilitated by a highly accomplished teacher. The research questions include: (a) To what extent is the knowledge building community productive, gauged based on communal and individual knowledge advancement? (b) What are the essential community interactions that sustain productive knowledge building? (c) What role does the teacher play in support of the above community dynamics?

## Method

### Participants

The participants were a class of 22 fourth-graders and their teacher at an elementary school in downtown Toronto. This study analyzes their knowledge building work in optics conducted over a three-month period supported by Knowledge Forum. The students had been using Knowledge Forum to conduct knowledge building since Grade 1. The teacher had accumulated strong expertise in facilitating knowledge building. An earlier study analyzed his three-year design experiment on improving classroom designs for knowledge building in optics, with significant improvement found each year (Zhang et al., 2009).

### The Knowledge Building Implementation

The optics inquiry was conducted in line with the knowledge building principles (Scardamalia, 2002). The teacher particularly focused on enhancing student collective responsibility for advancing community knowledge. Instead of having students work in fixed small groups, the teacher adopted an opportunistic collaboration design that encouraged students to define and elaborate progressive inquiry goals and, accordingly, adapt their participation structures over time (Zhang et al., 2009). As a result, the knowledge building process involved a dynamic flow between individual inquiry, small group work, and whole-class conversations and allowed students to group and re-group spontaneously based on their evolving goals. The students engaged in knowledge building discourse in Knowledge Forum to contribute and improve their ideas, mirroring and extending their discourse in the classroom. Their idea improvement was further supported by constructive use of authoritative sources (e.g., books) and experimental work. Student ideas, problems, data, and reading information were contributed to Knowledge Forum for sustained online discourse.

### Data Analyses

To analyze the growth of the community's knowledge space, we traced the online discourse in the seven views (workspaces) to identify progressive questions and ideas and, then, compared the questions and ideas against the curriculum guidelines. Individual knowledge growth was further assessed based on their portfolio notes. Each student wrote three portfolio notes online that summarized their understanding of light in the first, second, and third month of the inquiry, respectively. The analysis first looked at whether a portfolio note addressed each of the eight focal knowledge goals identified by the community (e.g. how lenses work). Following content analysis (Chi, 1997), student writing related to each goal was then coded based on two four-point scales (Zhang et al., 2007): (a) scientific sophistication (1 - pre-scientific, 2 - hybrid, 3 - basically scientific, and 4 - scientific) and (b) epistemic complexity (1 - unelaborated facts, 2 – elaborated facts, 3 – unelaborated explanations, and 4 - elaborated explanations), measuring effort to produce not only descriptions of the material world but theoretical explanations of hidden mechanisms (Salmon, 1984). Two raters independently coded 12 portfolio notes to assess inter-rater reliability, with full agreement for the coding of knowledge goals/themes, *Cohen's Kappa* = .83 for scientific sophistication, *Cohen's Kappa* = .75 for epistemic complexity.

We analyzed classroom video transcriptions, online discourse, and the teacher's reflection journal to identify community interactions and teacher scaffolding that had sustained productive knowledge building. The analysis integrated multiple levels and timescales, with holistic analysis of the whole initiative and detailed analysis of activity components informing each other (Lemke, 2000). In the first phase, the focus was on the community's three-month inquiry, as a whole, to understand how it started and evolved and its changing conceptual landscape. In the second phase, the focus was on each videotaped classroom episode, such as a whole-class conversation, a small group activity, or a computer lab session, to identify its focus, context, storyline, and connections with other episodes. In the final phase of detailed analysis, the focus was on each conversation turn in an episode to understand the nature of the discourse move in relation to its preceding and following conversation as well as the storyline of the episode identified in Phase 2. A more narrow approach to conversation analysis (Sawyer, 2006) was used to identify patterns through an emergent inductive process (Strauss & Corbin, 1998) without applying a predefined coding scheme and counting each code. Specifically, we read and re-read the transcriptions and coded specific moves (e.g., asking a challenging question about a peer's idea, providing evidence) represented by different conversation turns, with the speaker of each turn identified as either a student or the teacher. We then searched for connections across the specific codes in light of the contexts of the episodes and the evolution of the whole inquiry, and aggregated the codes into fewer, more encompassing themes, representing community-level (system-level) processes sustaining knowledge building. Each theme involved multiple sub-themes, with the encompassed specific moves of the teacher and those of the students considered side by side. The initial themes and sub-themes were then refined, elaborated, and validated through theme to theme, theme to data, data to data comparison, including triangulating the identified themes and sub-themes to the analysis of the teachers' reflection journals and the online discourse.

## Results

### To What Extent is the Knowledge Building Community Productive?

Tracing the online discourse helped to understand the conceptual scope and depth of the community knowledge space. Over the three months, a total of 168 notes (excluding the 66 portfolio notes) were created by the community in several views in Knowledge Forum. In each view, the students identified deeper questions leading to progress in understanding. For example, in the view of "Colors of Light and Rainbows," the students progressively examined how rainbows are created, why the colors are always in the same order, primary and secondary colors, the nature of white light, and how we see colored objects. The student discourse addressed all the expectations for Grade 4 in the Ontario Curriculum and further led the students to understanding many issues expected for Grades 6 and 8, such as light waves, color vision, colors of opaque objects, concave and convex lenses, the law of reflection. Individual knowledge advancement was assessed based on content analysis of student portfolio notes written in the three months, respectively (Table 1). Repeated measures ANOVAs revealed significant growth across the three months in the number of focal issues/goals addressed ($F$ (2, 42) = 43.03, $p < .001$, partial $\eta^2 = .67$), epistemic complexity ($F$ (2, 42) = 69.20, $p < .001$, partial $\eta^2 = .77$), and scientific sophistication ($F$ (2, 42) = 70.60, $p < .001$, partial $\eta^2 = .77$).

Table 1: Student knowledge advancement measured based on their portfolio notes.

| | 1st Month | 2nd Month | 3rd Month |
|---|---|---|---|
| # of focal issues addressed so far | 4.27 (1.83) | 6.41 (2.02) | 7.36 (.95) |
| scientific sophistication (1- pre-scientific to 4 - scientific) | 1.48 (.66) | 2.25 (.74) | 2.81 (.48) |
| complexity (1-unelaborated fact to 4-elaborated explanation) | 1.32 (.64) | 2.04 (.78) | 2.91 (.70) |

*Note*. Numbers are means and standard deviations.

### Community Dynamics and the Role of the Teacher

We analyzed the video transcriptions, teacher reflection, and student online discourse to understand how the productive knowledge building was achieved. A specific focus was on identifying significant moments of progress and then tracing backward and forward across activity contexts to understand how the process came about and where it further went. For example, a whole class conversation was conducted on May 16 that led progress in understanding how light interacts with opaque materials. The conversation began with students sitting in a circle and sharing observations of different materials. The teacher then highlighted a question originally raised by a student in Knowledge Forum: "Are all opaque things reflective like mirrors? " The following shows the first four of the 30 minutes or so discussion to address this and related problems.

Table 2: A Section of the Transcriptions of the May 16 Whole Class Conversation.

| Transcription | Analysis highlight |
|---|---|
| 2:20 GM: Well, there'[re] bricks, which are still opaque. But they're not reflective. But I don't know what they are called, like that kind of opaque. JL | Identifying non-reflective opaque materials, using a tentative voice. |
| 2:31 JL: I think all opaque materials are reflective, except not all of them reflect light back. … OK, let's just say um like…a yellow carpet… your eyes would be able to see the yellow of it because it would only reflect yellow light. That means like that sort of like a tissue for example that would only reflect white, except the yellow carpet, since it's like green mixed with red, I believe. Then the beam of red [and green] light would touch us and your eyes would take it in as yellow. | Contributing an alternative idea: All opaque materials are reflective, by analyzing a thought experiment (yellow carpet), drawing on knowledge of primary and secondary colors. |
| 3:58 Teacher: So you're saying everything is reflective then. Every opaque object is reflective to some degree. Oh, I hear some people disagree. Can you pass it on? [JL: SG.] | Revoicing student idea; highlighting contrasting perspectives. |
| 4:07 SG: What about wood? Wood isn't reflective. JL. | Bringing in an anomaly. |
| … | |
| 4:53 FJ: I think if wood is shiny and polished, you could see your reflection. I think it's mostly just shiny objects so it depends on what kind of wood you have, what kind of table you have, if you see your reflection. SG. | Re-analyzing and interpreting the instance as non-anomalous. |
| 5:12 SG: Like if you had a glass table. | Supporting fact. |
| 5:16 Teacher: The question is: Are all opaque objects reflective? Have we answered that? … Do all opaque objects reflect light? Anyone has a theory or evidence to support that? So, SG, it's yours to pass. [SG: DN.] | Highlighting/reminding a focal problem and promoting reflection on progress. |
| 5:35 DN: Um, actually all opaque objects do reflect light, because they reflect their own color. So we see them as whatever color they are. TS. [inaudible student talking] | Articulating an idea and its supporting thoughts. |
| 5:57 Teacher: Hold on, let's hear him talk. | Maintaining conversation norms. |
| 5:59 TS: If they didn't reflect their own color, you wouldn't see a brick of red, or someone's t-shirt as purple or whatever. RP. | Extending and elaborating idea. |
| 6:11 RP: What about black? | Bringing in an anomaly. |
| 6:14 Teacher: Don't throw it back to him. Give your theory. | Maintaining norms; encourage initial thoughts. |
| 6:18 RP: I don't think black reflects. I think that black might reflect light, but it might not. Because we had a reading today that um all the colors of the rainbow make white light and there is a note in the database about that, and everything reflects its own color. But it didn't say anything about black. EY. | Summarizing a reading and an online note and identifying black as an unaddressed issue. |

A number of important ideas emerged from the conversation (e.g., all opaque things are reflective, expect black ones). Tracing each of these ideas back into the prior inquiry activities revealed striking historical connections and incremental moves. For example, the idea that all opaque things are reflective is rooted in a small group reading and related experiments on transparent, translucent, and opaque materials conducted on May 9; the notion that opaque objects reflect light of their own colors traced back to May 12 when group readings and discussions occurred focusing on how people see colored objects. These ideas were circled back into the current conversation leading to new inferences, connections, and meanings and further informing deeper problems and ideas.

Analyzing community interactions with each episode, such as the one shown above, and searching for cross-episode connections helped to identify essential processes/patterns enabling collaborative productivity in the community. Four thematic processes emerged, in consistence with the self-organization mechanisms sustaining knowledge creation synthesized in the beginning of this paper.

## Process 1: Accumulating a Highly Variable Stock of Information and Ideas and Mobilizing Information Flow and Connection

*Sub-process 1a: Accumulating a communal knowledge base that involves a richly connected history of ideas, with new inquiries evolving out of the old.* The students accumulated a communal knowledge base—supported by Knowledge Forum—that helped to enable historical continuity between the old, current, and future work. Rich information flow took place on the basis of the community's knowledge base. Specifically, the students shared their understanding and challenges so as to construct a common ground while shaping emergent ideas and goals. Old ideas, problems, and inquiry works were constantly referred to and circled back to advance the current inquiry and discourse and further inform emergent goals (e.g., "going back to what X said about… , so…"). There was rich cross-referencing between the online and offline spaces and between different inquiry activities, as illustrated through the May 16 conversation analyzed above.

The teacher modeled making connections across ideas and activities and attempted to anchor all his major input in student history of inquiry. Specifically, he worked with the students to formulate emerging inquiry focuses/themes in the contexts of student prior inquiry and current interesting events. He helped to ground classroom conversations in relation to students' accumulated work and ideas, and highlighted connections between current inquiry and prior readings, experiments, and discussions. He also facilitated the online-offline connections, with important ideas and related inquiry work recorded in Knowledge Forum for further discussion and inquiry (e.g. "Is anyone writing a note about this experiment?"). When multiple small groups were conducting different investigations in the classroom, the teacher was often called upon by the students, who wanted to share with him their wonderments and excitements. He captured important progress of understanding and then catalyzed community wide discussion, online and offline.

*Sub-process 1b: Making constructive use of information from reading.* Efforts to accumulate a vibrant knowledge base were advanced through student constructive use of reading. To address important problems emerged from their inquiry, the students found and read a large amount of material. They worked in groups to understand difficult text, using reading strategies (e.g., questioning, reviewing, summarizing) to deepen their comprehension and discussion. Reading professional text and bringing important concepts and information to the community discourse helped the students to expand the knowledge base they could work with and appropriate sophisticated language to represent and process their ideas. It further engaged the students in reflective dialogues between their local understanding and knowledge out in the world (e.g., science communities). Consistent findings from the readings were synthesized and used by the students to support and extend their ideas. Inconsistencies were identified leading to further idea development.

The teacher modeled monitoring knowledge gaps in the community's knowledge space and introduced new readings by talking about existing questions and related ideas. As the inquiry went deep into the domain, many of the students-generated problems required readings above their current grade level. The teacher communicated his deep trust and high expectation and inspired students to collaboratively deal with difficult text for deep understanding. He consulted student interest in the readings to set up reading groups. Occasionally, he participated in the small groups to co-analyze and reflect on key information from readings, design related experiments, model various reading strategies. Across the inquiry activities, he helped the students to see connections between the current work and what they had read.

## Process 2: Sustained Idea Development, Powered by Positive Feedback Loops

*Sub-process 2a: The more we know, the more we can learn and generate: incremental idea development.* The students connected to and made use of their shared knowledge base and peers' intellectual input to generate new ideas and deepen understanding. In the online and offline discourse (e.g., Table 2), the students made connections to and drew upon existing ideas and inquiry work to generate new ideas to address the focal problems. They often contributed ideas in the form of tentative conditional statements (e.g., maybe…) open to critics and further input (Engestrom, 2008). Their peers then responded to extend and elaborate these ideas; to contribute related thought experiments (e.g., what will happen if…), analogies (e.g., between eye and camera), observations, and supporting facts; to present alternative ideas and anomalies; to identify subordinate questions and distill challenges; and to summarize different perspectives for deeper conceptualizations. These interactive moves are characteristic of distributed reasoning observed in productive research teams (Dunbar, 1997), with existing ideas and knowledge operations constantly taken up by peers to enable further advancement.

The teacher nurtured among his students a sense of epistemic agency and empowerment. He communicated his deep trust that everyone had something worth saying and could contribute to knowledge advancement. In classroom conversations, he (a) engaged in active listening to and reading of student ideas; (b) expressed encouragement, interest, and excitement; (c) asked questions on student ideas for clarification and deeper thinking; (d) highlighted interesting ideas (including misconceptions) and questions, along with possible connections, controversies, and gaps; and (e) re-voiced student ideas (see Table 2) to make them more explicit and formal in light of related domain concepts (Hmelo-Silver & Barrows, 2008; O'Connor & Michaels, 1992).

*Sub-process 2b: The more we know, the more we realize that we do not know, driving us to know more: Progressively identifying problems and formulating deeper goals.* The students actively monitored and distilled challenging questions emerging from their discussions, reading, and experiments. Integral to their classroom conversations, they participated in meta-discourse to identify emergent questions and review conceptual connections among the questions so as to evolve interconnected deeper goals. For example in the May 16 conversation, the students identified questions such as: Does light reflect off of black opaque objects? How comes a mirror reflect light of all colors? Identifying these issues led to further idea development in the subsequent discourse, which then brought further problems of understanding to the fore.

The teacher explicitly encouraged students to take on high-level responsibilities and knowledge operations for sustained deepening of understanding. He encouraged students to ask 'why, why,' and not to be content with a superficial answer. He built on student discussions of emergent questions to propose his framing of focal inquiry goals, which was then discussed by the community. He engaged in deep listening to and reading of student ideas and questions, highlighted important problems emerged from classroom conversations, and asked questions to deepen student-generated ideas instead of directing the students to new tasks (Zhang et al., 2009). He also reminded/brought back problems that had not been addressed and requested reflective review of progress (e.g., "Could someone summarize what we've come to so far at this talk?"). The students oftentimes did the same supported by the teacher's modeling.

## Process 3: Critical Examination and Selection of Ideas, as Negative Loops

*Sub-process 3a: Individual reflection and critical idea examination in the discourse*. Individual reflection appeared to function as a bottom-level mechanism of idea examination and selection to make sure that everyone contributed clearly presented and carefully reasoned out ideas addressing communal goals. When presenting information and ideas, the students acknowledged the sources and connections to prior work of peers. They used conditional statements and indicated both what they knew and what they were not sure about, in reflection of the "half-baked" nature of their ideas. Student idea input was further examined and selected through critical dialogues. In the dialogues (e.g., Table 2), the students challenged peer ideas by presenting alternative explanations, identified and analyzed anomalies, and raised questions that challenged existing explanations. The competing explanations, possible anomalies, and challenging questions *per se* then became objects of critical discourse. The examination of two competing explanations might lead to giving up one of them; but in many other cases, it led to rising above different perspectives towards more complicated explanations.

Related to idea examination and selection, the teacher was cautious not to be the judge of student ideas. But rather, he highlighted rules of reflective contributions (e.g., contribute ideas with details) and engaged his students in meta-discourse to review ideas, monitor gaps, conflicts, and challenges, and reflect on progress, using identified contrasting perspectives to stimulate deeper examination and analysis.

*Sub-process 3b: Empirical testing of ideas.* Focusing on important questions and ideas about light, the students designed and conducted experiments and observations, often in small groups. The students collaboratively identified focal problems and deeper questions to be addressed, generated theories, designed experiments, interpreted findings, and discussed new insights. They brought their new understanding and supporting evidence to the subsequent discourse for broad sharing, collective scrutiny, and further build-on. Their experiments were emergent and idea-centered instead of as pre-scripted tasks, with the goal of collecting data to examine and develop explanations.

The teacher listened to and read student ideas and suggested the need of evidence to examine different ideas, such as by saying: "I'm interested in your theory. Can you design an experiment to test this idea?" He listened to student proposals for experiments and promoted reflective thinking, such as by asking: "What are the questions you are trying to answer?" "What are the steps you will go through?" Based on the students' proposals, the teacher helped them to find needed materials and instruments.

## Process 4: Distributed Emergent Control

*Sub-process 4a: Co-constructing principles, strategies, and support structures*. In all types of inquiry work, the students reflected on and talked about how things should be done, and why, leading to deep understanding of what it means to work as a knowledge building community. They talked about collective ownership over inquiry questions and ideas, so that a community member might read and find information that was beyond his/her immediate personal focus. They talked about norms of knowledge building discourse, such as active listening/reading, contributing ideas in full paragraphs with evidence and details, and making connected and non-redundant input (e.g., do not write a note that simply repeats a question or idea). The community members also co-developed and utilized a variety of externalized structures to guide, assist, and deepen knowledge building. For example, to represent knowledge goals and guide their writing of reflective portfolio notes, the students collectively generated a list of thematic questions (e.g., how mirrors work) in line with their focal goals and then turned the questions into a set of new scaffold labels in Knowledge Forum. To assist experimental design and reporting, the teacher and his students discussed and agreed on key elements of a scientific experiment (e.g., theory, steps, results). The elements were then listed by each student in the front page of his/her laboratory notebook.

In the classroom discourse/meta-discourse, the teacher discussed, modeled, and reminded students of the basic conversation norms and rules, which encouraged collective engagement, reflective thinking, and sustained improvement of ideas. He shared with the students his classroom design ideas open to their input.

*Sub-process 4b: Collective responsibility, adaptive roles, and opportunistic planning.* The students monitored progress of understanding, identified challenging issues to be addressed, and synthesized important knowledge advances in each Knowledge Forum view. They brought the advances and challenges to whole-class conversations to look into possible deep connections and formulate further inquiry goals. Individuals then decided which aspects of the inquiry goals they wanted to contribute to and in what ways (e.g., theorizing, experimenting, reading), with those who had shared interest often forming into temporary small groups. The community awareness of who was working on what further created a social pressure that helped to increase individual accountability (Zhang et al., 2009). Sometimes, student peers would ask a student/group: "You have been working on this for a long time. You need to tell us what you have found."

The teacher's classroom design was characteristic of opportunistic planning (Sawyer, 2007). He identified related "big ideas" in the domain and thought about how the ideas might be approached in light of his understanding of knowledge building principles (Scardamalia, 2002) and observations of previous classes. Doing so helped him to create a big picture in mind about how the inquiry might evolve while leaving all detailed processes open and co-improvised with his students (Zhang et al., 2009). He actively observed student work and listened to their ideas to understand their progress and challenges; contributed to meta-discourse to formulate, distill, and highlight knowledge goals; created, linked, and adjusted view structures in Knowledge Forum in line with the evolving goals; listened to student inquiry plans and gave suggestions; and walked between small groups to understand their progress and offer on-site advising. Sometimes, he also proposed action plans. When doing so, he always connected his proposals to existing ideas, questions, and input from the students, such as by saying "I'm interested what X said earlier about…"

## Discussion

The content analysis of the students' online discourse and portfolio notes revealed high-level collaborative productivity achieved by the community. The qualitative analyses of the classroom videos and teacher reflections further elaborated four interrelated processes that sustained the collaborative productivity, with the teacher's roles identified in each of the processes.

Efforts to engage students in distributed, opportunistic collaboration for knowledge building often face the questioning of how students possibly know what to do to make productive progress. Such questioning is rooted in a common belief that students need a "strong leader" to chart and organize the process of knowledge building and tell them what tasks/sub-tasks to be done and in which ways (e.g., using scripts). This study elaborated collaborative productivity as emergent, self-organizing processes. Neither the teacher nor the students knew beforehand how the inquiry would exactly unfold. The course of inquiry and collaboration and the key moments of progress emerged from an interactional process in which the teacher and his students co-contributed to the unfolding classroom flow. Student interactions augmented through the shared knowledge space in Knowledge Forum enabled continuous and incremental idea development. They identified important and relevant ideas from the past as the stepping-stones of their new inquiry, triggering deeper ideas and problems. Each major idea was embedded in the evolving intellectual history of the community, gaining support from the past work and further informing and enabling future idea development (Tabak, 2004). As a result, new deepening goals and plans emerged as the inquiry history unfolded. Instead of having the teacher make high-level decisions regarding the rules, structures, goals, and procedures, the community members took on collective responsibility for evolving goals and developing productive practices and structures.

The significant advances in the community's knowledge did not come about through sudden insights that departed from existing knowledge, but through historical build-on, incremental refinement, and accumulative selection of ideas (Cziko, 1995; Dunbar, 1997; Sawyer, 2007). Student ideas and questions were constantly taken up by peers and used as the input to further operations of knowledge, leading to idea generation, elaboration, diversification, and improvement. The advanced understanding further helped to inform deeper problems at the frontier, enabling sustained cycles of progressive problem solving. Thus, the more they knew, the more they could generate; and the more they realized that they needed to know. These positive loops fueling sustained idea generation were coupled with critical idea examination and selection. In the knowledge building discourse, ideas were often contributed in a tentative voice open to critics and further input; competing explanations, possible anomalies, and challenging questions were raised calling for further examination of ideas; empirical evidence was collected and brought to discussions.

Elaborating the self-organization mechanisms underpinning collaborative productivity sheds light on various aspects of community scaffolding—the community as the provider as well as recipient of scaffolding—through emergent, distributed processes (Davis & Miyake, 2004; Tabak, 2004). Members use their historically accumulated ideas to support the current work and inform deepening goals; New ideas and inquiry strategies are objectified and selectively retained to enabled further advances; Collaborative discourse among the members enables sustained chains of distributed reasoning (Dunbar, 1997); Supportive principles, rules, and external scaffolding structures are developed, monitored, and adapted through reflective discourse. The roles of the teacher aligned with each of the processes exemplify specific teaching designs and strategies to nurture and support knowledge building. Viewing these specific roles and strategies through the lens of self-organization helps to understand how these roles synergize with one another (Tabak, 2004) and how teacher scaffolding leverages community scaffolding for productive knowledge building.

Future work needs to further examine the self-organization mechanisms that sustain collaborative knowledge building in different grade levels and explore if some of the processes (e.g., positive and negative loops) can be used as leverage points to help transform classrooms into creative communities.

## References


Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Erlbaum.

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves.* Chicago, IL: Open Court.

Barron, B., Darling-Hammond, L. (2008). How can we teach for meaningful learning? In L. Darling-Hammond, B. Barron, D. Pearson, A. H. Schoenfeld, E. Stage, et al. (Ed.), *Powerful learning* (pp. 11-70). San Francisco: Jossey-Bass.

Chatzkel, J. L. (2003). *Knowledge capital*. New York: Oxford University Press.

Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of the Learning Sciences, 6,* 271-315.

Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 313-335). Cambridge, UK: Cambridge University Press.

Cziko, G. (1995). *Without miracles: Universal selection theory and the second Darwinian revolution*. Cambridge, MA: MIT Press.

Davis, E. A., & Miyake, N. (2004). Explorations of scaffolding in complex classroom systems. *Journal of the Learning Sciences, 13*, 265-272.

Dunbar, K. (1997). How scientists think: Online creativity and conceptual change in science. In T. B. Ward, S. M. Smith & S. Vaid (Eds.), *Conceptual structures and processes: Emergence, discovery and change* (pp. 461-493). Washington, DC: APA Press.

Engeström, Y. (2008). *From teams to knots*. New York: Cambridge University Press.

Hakkarainen, K. (2009). A knowledge-practice perspective on technology-mediated learning. *International Journal of Computer-Supported Collaborative Learning, 4*, 213–231

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction, 26*, 48-94.

Lemke, J. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture, and Activity, 7,* 273-290.

Meadows, D. (1999). *Leverage points: Places to intervene in a system*. Hartland, VT: The Sustainability Institute.

Miyake, N., & Norman, D. A. (1979). To ask a question, one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior, 18*, 357-364.

O'Connor, M. C., & Michaels, S. (1992). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly, 24,* 318-335.

Salmon, W. C. (1984). *Scientific explanations and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Sawyer, R. K. (2006). Analyzing collaborative discourse. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 187-204). New York: Cambridge University Press.

Sawyer, R. K. (2007). *Group genius: The creative power of collaboration*. New York: Basic Books.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Eds.), *Liberal education in a knowledge society* (pp. 67-98). Chicago, IL: Open Court.

Scardamalia, M. (2004). CSILE/Knowledge Forum®. In A. Kovalchick, & K. Dawson (Eds.), *Education and technology: An encyclopedia* (pp. 183-192). Santa Barbara, CA: ABC-CLIO, Inc.

Sternberg, R. J. (2003). The development of creativity as a decision-making process. In R. K. Sawyer et al. (Eds.), *Creativity and development* (pp.91-138). New York: Oxford University Press.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed). Newbury Park, CA: Sage.

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist, 42*, 115-121.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences, 13,* 305-336.

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in 9- and 10-year-olds. *Educational Technology Research and Development, 55,* 117–145.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge building communities. *Journal of the Learning Sciences, 18*, 7–44.

# Extending the Self-Explanation Effect to Second Language Grammar Learning

Ruth Wylie, Kenneth Koedinger, Teruko Mitamura
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, Pennsylvania 15217
Email: rwylie@cs.cmu.edu, koedinger@cmu.edu, teruko@cs.cmu.edu

**Abstract:** Self-explanation is an instructional strategy that has shown to be beneficial for math and science learning. However, it remains an open question whether these benefits will extend to other domains like second language grammar learning. Within the domain of the English article system (teaching students when to use *a*, *an*, *the*, or no article at all), we compare two computer-based tutoring conditions in an in vivo classroom study. In the article choice condition, students select the correct article to complete the sentence. In the explanation choice condition, students are given a sentence with the correct article highlighted and choose the rule or feature that best explains the article use. Students (N=101) in both conditions show significant learning on both procedural (article choice) and declarative (explanation choice) tasks. Not surprisingly, we found that declarative instruction (explanation choice) led to significant learning of explanations, while procedural practice (article choice) led to significant learning of the procedures. More interestingly, we also found evidence of cross-type transfer such that declarative practice led to procedural gains and procedural practice led to better understanding of the declarative rules. In general the effects of prompted self-explanation appeared somewhat stronger than those of procedural practice.

## Introduction

Self-explanation is an instructional strategy that has lead to increased learning by encouraging students to focus on key features of the material (Roy & Chi, 2005). However, the majority of this work has been done in math and science domains like physics (Chi, et al., 1989; Conati &VanLehn, 2000), geometry (Aleven & Koedinger, 2002), and biology (Chi, et al., 1994), with little existing work in domains like second language grammar learning. The first self-explanation studies were correlational studies that demonstrated that students who self-explain more learn more (Chi, 1989). Follow-up studies have shown that prompting students to self-explain increases learning (Chi, et al., 1994), and that having students select explanations from a list of options (rather than constructing free-formed responses like in previous studies) is also a successful technique (Aleven & Koedinger, 2002; Renkl, 1999). However, despite being called a domain general strategy (Roy & Chi, 2005), the self-explanation effect has not been extensively tested in domains other than math and science with the exception of McNamara's (2004) work on reading comprehension. Our work differs from previous work in its focus on second language grammar acquisition, a problem-solving domain where there is a strong intuition that immersive practice, without reflection on rules, is more natural and perhaps best.

In this paper we compare math and science learning and second language grammar acquisition and discuss why self-explanation may or may not be beneficial for learning English grammar constructs. We then describe two systems that we built to evaluate the self-explanation effect in an empirical classroom-based study within the context of learning the English article system (teaching students when to say "a pencil" versus "the pencil"). As expected, students in both conditions demonstrate significant learning gains on their tutored skill: students in the tutored-practice (article choice) condition learn how to select the correct article, and students in the self-explanation condition learn how to select correct explanations for article use. More surprisingly, students in both conditions also show cross-type transfer with students in the article choice condition learning how to select explanations and students in the self-explanation condition learning how to select the correct article. These results are promising and suggest that self-explanation is beneficial for second language grammar acquisition and leads to the refinement of both procedural and declarative knowledge.

### Article Domain

We chose to focus on the domain of non-native English speakers learning the English article system because of the importance of acquiring article use skills, the rule-based structure of the article system, and second language acquisition theory that suggests articles are particularly well suited for a rule-based instructional approach. Grammar errors can cause writing to become "intelligible", "irritating", or both (Ellis, 1994), and articles are one of the last grammar points for English language learners to acquire (Master, 1997). We also chose article use as a domain because its rule-based nature makes it a good candidate for studying the effects of self-explanation in language learning. In contrast to domains like vocabulary learning where students must memorize arbitrary mappings, article usage is generally determined through a set of heuristics based on features of the sentence (although many exceptions exist). For example, in the sentence, "Yesterday, I bought a car. Today, *the* car broke," the article *the* is used because the noun *car* has already been mentioned. Furthermore, according to second language acquisition theory, some grammar constructs are better suited for rule-focused instruction than others. For example, articles are a type of grammar construct that is often not required for successful communication (i.e. readers likely understand the sentence, "Yesterday, I went to (no article) store" even though "Yesterday, I went to *the* store" is correct), and thus articles and their function within a sentence are unlikely to be noticed by learners without formal, rule-focused instruction (Williams & Evans 1998).

## Domain Differences

While self-explanation has been very successful at increasing learning in math and science domains, there are inherent differences between math and science and second language learning which may affect the success of self-explanation in this new domain. One key difference lies in the pedagogical goals of the domains. In math and science, often students are expected to be able to solve problems *and* explain the underlying principles. A typical example of this expectation is in geometry where students are asked to determine the angles of a triangle and provide the reason or rule(s) for their answer. However, the marker of a successful language learner is fluency, and knowing when and how to use a particular construct is much more important than knowing why. In fact, most native speakers of a language have no explicit knowledge of the rules driving their article decisions yet rarely make mistakes. Another difference lies in the presence and absence of exceptions to rules. In math and science, there are no exceptions. Provided that the proper conditions are met, a given rule will always apply. However, there are frequent exceptions to grammar rules. For example, one rule listed in an English as a Second Language grammar book states that no article should be used before the name of a disease (Cole, 2000); for example, *He has (no article) diabetes.* However, many exceptions exists (e.g. *He has* ***the*** *flu*, *She has* ***a*** *cold.*) Perhaps encouraging students to focus on rules and features of the sentence, which may sometimes prove to be unreliable, is not an effective learning strategy.

## Tutor Designs

In our study, we compared a condition where students were tutored on which article to use in a sentence (*article choice*) to a condition where students were tutored on the features of the sentence relevant to article use (*explanation choice*). Both tutors were developed using the Cognitive Tutoring Authoring Tools (Koedinger, et al., 2004) and deployed online using Java WebStart. All student actions (answer selections, hint requests) were logged and time-stamped. Both tutors had 56 example sentences and addressed eight rules for making article decisions (e.g. "If the noun has already been mentioned, use *the*" or "If the noun is general and plural, use *no article*."). In order to finish working with the tutor, students must have answered all questions correctly.

The article choice condition (Figure 1) mimics cloze or fill-in-the-blank activities found in many second language learning textbooks. Using dropdown menus, students select the article ("a", "an", "the", or "no article") that best completes the sentence. They receive immediate feedback on their selection and have access to a series of on-demand hints. This form of instruction is designed to give students practice using articles but does not require them to give a reason or explanation for their choice.

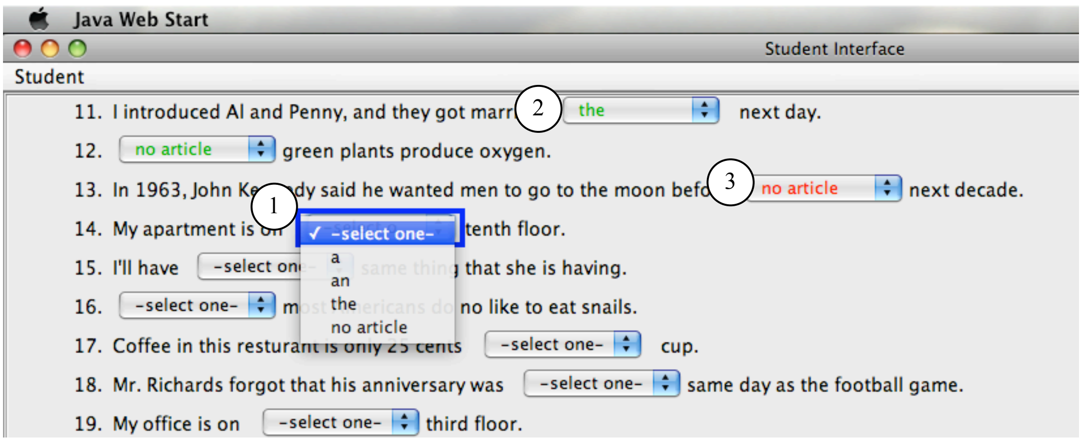



Figure 1. In the article choice condition, students choose the article (a, an, the, no article) that best completes the sentence (1), and receive immediate feedback on their selection. If the answer is right, it turns green (2), and red if it is wrong (3).

In this study, we chose to operationalize self-explanation in the form of a menu-based explanation choice tutor. In this condition (Figure 2), students are presented with a sentence with the target article highlighted and asked to choose the rule that best explains the article use. Again, students receive immediate feedback on their selection and have access to hints. In this condition, students see examples of correct article use but do not make any article decisions of their own.

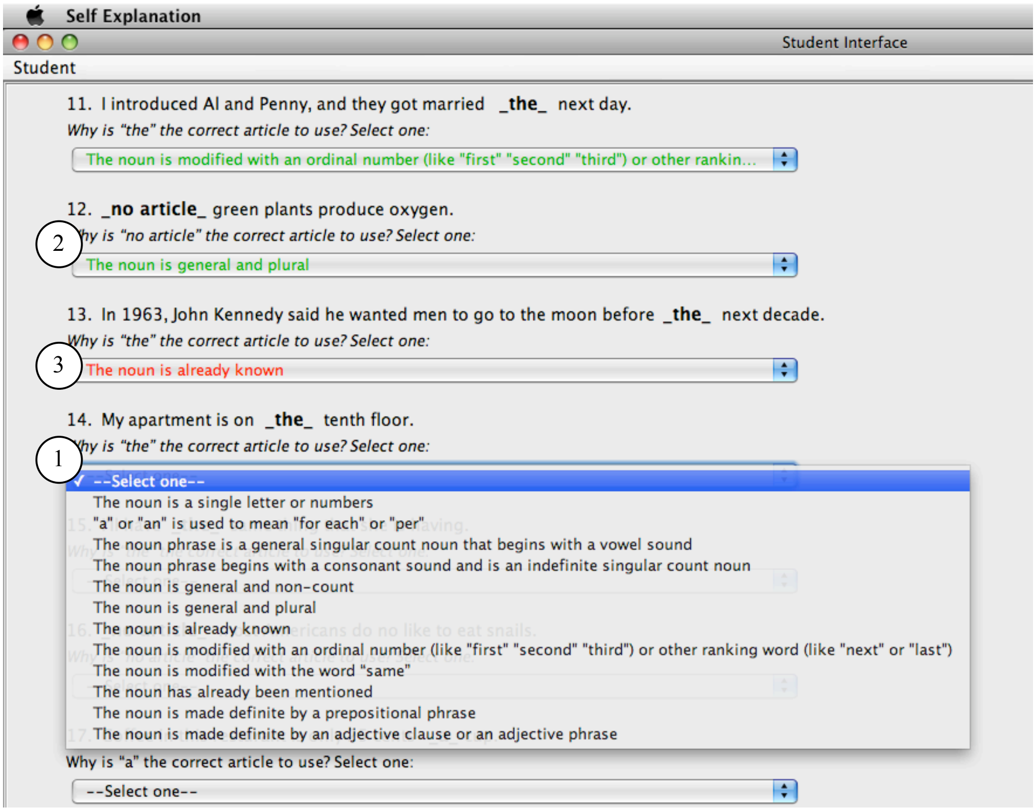



Figure 2. In the explanation choice condition, students choose the feature of the sentence that best explains the article use (1). Identical to the article choice condition, students receive immediate feedback on their selection. If the answer is right, it turns green (2), and red if it is wrong (3).

Both tutors employed a similar series of hints that were presented upon student request. The hints first identified the key feature(s) of the sentence, then provided the rule, and finally, in the article choice condition, told students which article to select (Table 1).

Table 1: Example hint sequence for the article choice and explanation choice tutors.

*Target Sentence: My office is on the third floor.*

| Article Choice Hints |
| --- |
| 1. "Third" is an ordinal number. |
| 2. Use "the" with ordinal numbers and other ranking words like "next" or "last". |
| 3. Please select "the" from the highlighted menu. |

| Explanation Choice Hints |
| --- |
| 1. "Third" is an ordinal number. |
| 2. Please select "The noun is an ordinal number (like "first", "second", "third") or other ranking word (like "next" or "last"). |

## Hypothesis

Our study addresses the question of whether self-explanation is a helpful instructional strategy for second language grammar learning. We hypothesize that students in the explanation choice condition will show greater learning gains on article choice and explanation choice tasks than those in the article choice tutor. One hypothesis for how prompts for self-explanation enhance learning is that they encourage students to notice relevant features of the problem and enable students to become aware of gaps in their own knowledge (Roy & Chi, 2005). This feature focusing technique may be beneficial to second language learning students by helping them attend to parts of the sentence that are important for making article decisions and ignore irrelevant parts.

Within the field of second language learning, there have been several studies that have investigated the differences between explicit and implicit grammar instruction. The findings from this study also contribute to this debate, since although both conditions are examples of focus on form instruction (Doughty, 2001), the explanation choice condition which supports explicit practice of rules is an example of explicit instruction, while the article choice condition affords a more implicit approach to learning. While both approaches have their own merits, according to a meta-analysis by Norris and Ortega (2000) instructional strategies that include explicit focus on rules, like the explanation choice condition, are more effective than strategies that are more implicit in nature, like the article choice condition.

Self-explanation may also help students by strengthening multiple processes by which to make article decisions. For example, students could solve the problem through an implicit strategy in which they choose the article that sounds the most correct, or, if they become proficient at the rules, they could use an explicit strategy and make their article decision based on the features of the sentence and the rule that applies. Thus, even though the ultimate goal is to create expert *users* of the rule and not expert *explainers* of the rule, self-explanation may still be a beneficial instructional technique.

## Method

The study took place during one 50-minute class period. Students (N=118) were adult English language learners (mean age = 27.9, SD=7.2) enrolled in the University of Pittsburgh's English Language Institute who came from a variety of first language backgrounds. In total, there were 13 first languages represented; however the majority of students spoke Arabic (37%), Korean (18%), or Chinese (15%) as their native language. Students participated in the study as part of their normal grammar class. There were three class levels: intermediate (n=30), intermediate-advanced (n=61) and advanced (n=27). The session started with an introduction to the tutoring and testing interfaces as well as a brief overview of the rules covered by the tutor. As adult language learners, all of the students had received some prior instruction on article use, and the goal of the demonstration was simply to introduce the vocabulary that was used in the tutors and to explain the tasks that the students would be completing. Students then completed the article choice and explanation choice pretests. Next, students were randomly assigned within each class to either the article choice or explanation choice condition and completed the tutored problems. Finally, students did the article choice and explanation choice posttests (Figure 3). Pre- and posttest forms were counterbalanced.

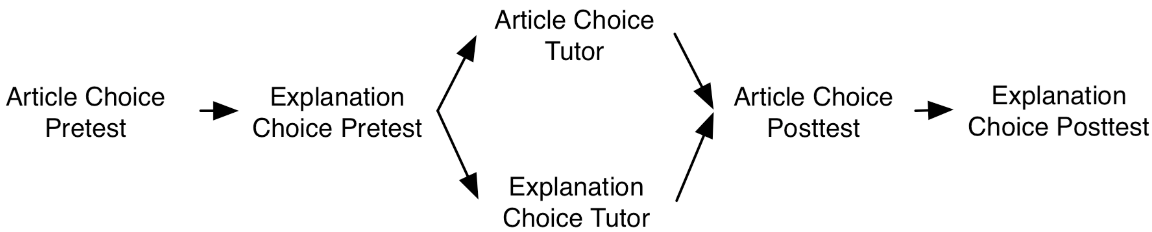



Figure 3. The sequence of tasks students completed as part of the study.

### Measures

All students were assessed on both procedural knowledge (article choice) and declarative knowledge (explanation choice) items. Article choice items were isomorphic to those in the article choice tutoring condition, and students chose the article that best completed the sentence. The explanation choice items were also identical in form to those in the explanation choice tutoring condition, where students were given a sentence with the correct article highlighted and asked to explain why that article was used. Two test forms were created, each with twelve article choice and twelve explanation choice items. For each problem type, there were eight questions that used rules taught in the tutor (tutored items) and four questions that used rules that were not taught in the tutor (control items). These items were included in order to measure effects other than those from the tutoring system (e.g. students becoming familiar with the interface or students becoming fatigued).

### Data Sample

Since students participated in the study as part of their normal class activity, there was limited time within which to collect data. Overall, 86% (101 out of 118) of the students completed all tasks; however, attrition between conditions was not the same with 95% (55 out of 58) of students in the article choice condition and 77% (46 out of 60) of students in the explanation choice condition completing all tasks ($\chi^2(1, N=118) = 7.9$, p = 0.005). Pretests scores for the remaining students were not significantly different ($t(1, 99) = 1.31$, p = 0.192) (Table 2), but the trend is the direction of concern (i.e., consistent with the hypothesis that low prior knowledge students were dropped from the explanation choice sample) with the explanation choice group having an average pretest score of 62.9% and the article choice group at 58.9%. Based on observation and anecdotal comments from the classroom teachers, students who ran out of time before completing the posttests fell into one of two categories: students with low prior knowledge and students with high prior knowledge who were very meticulous and conscientious in their choices. One student did not take the pretest due to technical error and was dropped from analysis.

Table 2: After attrition, the pretest scores were slightly higher than before attrition but there was no significant difference between conditions.

| | Total Sample | | Sample After Attrition | |
|---|---|---|---|---|
| Tutoring Condition | N | Pretest Mean (SD) | N | Pretest Mean (SD) |
| Article Choice | 58 | 0.583 (0.162) | 55 | 0.589 (0.159) |
| Explanation Choice | 59 | 0.604 (0.145) | 46 | 0.629 (0.137) |

Even though the pretest scores were not significantly different and we have reason to believe that attrition may also occur because of the diligence of good students, we thought it best to use a statistical analysis designed to address such a situation. Propensity score matching (PSM) is a technique that is often used in quasi-experimental designs to correct for bias. PSM is similar to other types of matching but it combines several variables into one score on which to cluster participants. Prior work has shown that using propensity scores to divide students into five groups and then using this subclassification as an additional variable often removes over 90% of the bias caused by each of the covariates (Rosenbaum & Rubin, 1984). In our study, we wanted to account for bias due to attrition and thus first calculated a propensity score (1 = highly likely to complete the study, 0 = very unlikely to complete the study) for each student by running a logistic regression with the binary variable of whether or not students completed the posttests as the dependent variable. Covariates used in the regression include: the log of the time spent on the article choice pretest, the log of the time spent on the explanation choice pretest, the log of the time spent using the tutor, article choice pretest score, explanation choice pretest score, and course level (intermediate, intermediate-advanced, advanced). The participants were then divided into groups based on propensity score and this value was used as a between-subjects variable in the all the analyses.

The main reason for grouping students using propensity score and not a single measure was that individually none of the other variables were highly correlated with completion rate. However the propensity score, which combines several measures, is highly correlated with whether or not a student would complete the posttests ($r(115)=0.84$, $p < 0.001$) and thus a better metric on which to group.

## Results

The following analyses were done with and without the PSM variable and resulted in the same conclusions. Because we believe they are less biased and thus a more accurate description of the findings, here we report the results that include the PSM variable.

### Learning Gains

Repeated measures ANOVA analyses show that students in both conditions demonstrate significant pre to posttest improvement on both the article choice measure ($F(1, 95) = 29.44$, $p < 0.001$) and the explanation choice measure ($F(1, 91) = 15.09$, $p < 0.001$). In order to determine if tutoring condition had an effect on posttest scores, a MANCOVA was calculated using article choice and explanation choice posttest scores as the dependent variables, propensity group and tutoring condition as the independent variables, and article choice and explanation choice pretest scores as covariates. This reveals a significant effect for condition ($F(2, 88) = 5.38$, $p = 0.006$). Finally, in order to understand the specific affects of the conditions, we ran univariate between-subjects tests for each test type. These showed that condition significantly affected explanation choice posttest scores ($p=0.023$) but not article choice posttest scores ($p=0.153$).

Figure 4 uses normalized gain (Hake, 1998) scores to illustrate these results: both groups are learning the skill on which they were tutored (article choice students show improvement on the article choice task, and explanation choice students show improvement on the explanation choice task). More surprisingly, both groups are also learning the transfer skill (article choice students show improvement on the explanation items, and explanation choice students show improvement on the article choice items). Furthermore, the effects of the self-explanation condition appear stronger than those of procedural practice. Despite having no tutored practice on article choice items, students in the explanation choice condition are showing equal gains on the article choice measure compared to students in the article choice condition and significantly higher gains on the explanation choice measure.

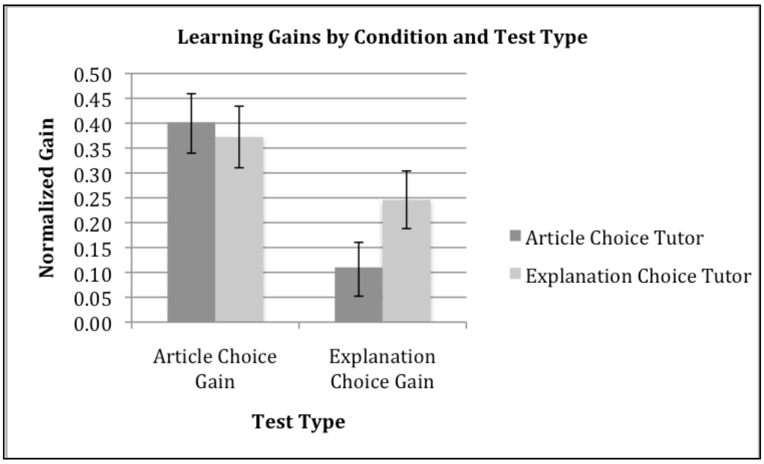
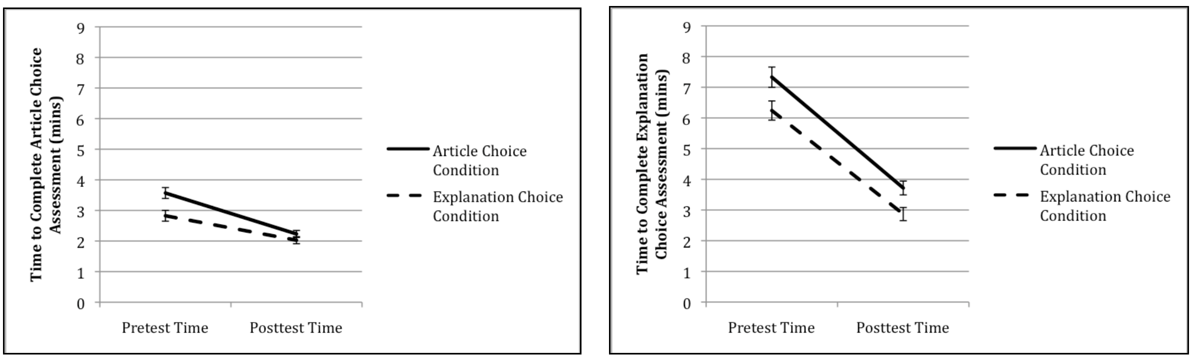



Figure 4: Normalized gains scores for article choice and explanation choice measures.

Surprisingly, we see cross-task transfer for both conditions: students using the article choice tutor show a small but significant gain on the explanation choice items, and students using the explanation choice tutor, who had no tutored practice with the article choice task, are showing equally high gain scores as those who used the article choice tutor. In addition to the tutored rules, students were also assessed on four control rules that were not included in the tutored material. These items were included in order to measure effects other than those from the tutoring system (e.g. students becoming more familiar with the interface, fatigue, etc.). For these untutored items, results of a repeated measures ANOVA investigating pre-to-post change across both conditions showed no improvement for the article choice items ($F(1, 95) = 0.01$, $p = 0.931$) and a significant decrease on the explanation choice items ($F(1, 91) = 7.59$, $p = 0.007$). These results support the conclusion that the observed learning gains on the tutored rules were the result of students' experiences with the tutors and not an unobserved, external factor. The decrease in explanation performance on untutored items may reflect students' tendency to use explanations consistent with the tutored items, which, of course, are incorrect for the untutored items.

## Processing Time

We also recorded the amount of time it took students to complete each assessment. Since measures of fluency include both accuracy and speed, reducing the amount of processing time that a student requires is an important goal. One concern with using self-explanation for teaching grammar is that it might take students longer to generate a response if they use an explicit, rule-based strategy than if they use an implicit one. A repeated measures ANOVA with log-transformed pre and posttest completion times as the dependent variables and condition and propensity group as the independent variables, show that students in both conditions are significantly faster at completing the posttest than the pretest for both measures (article choice: $F(1, 95) = 76.28$, $p < 0.001$, explanation choice: $F(1, 91) = 275.32$, $p < 0.001$) (Figures 5 & 6). A MANCOVA analysis with log-transformed completion times for article choice and explanation choice posttests as dependent variables, condition and propensity group as independent variables, and the log-transformed completion times for the pretests as covariates shows no main effect for condition ($F(2, 88) = 0.37$, $p = 0.689$). While students in both tutoring conditions are completing the posttests significantly faster than they complete the pretests, students in the self-explanation condition are completing the posttests just as fast as those in the article choice condition.




Figures 5 & 6: Students in both conditions and for both measures completed the posttest significantly faster than the pretest. More importantly, self-explanation instruction does not appear to be hinder students' speed in completing the mesures.

## Hint Usage

As mentioned above, the hint structure of the two tutors was very similar, and thus if students in both groups heavily relied on the hints, their experiences with the tutors might be similar despite the task (article choice vs explanation choice) differences. However, the log data show that students rarely request a hint (Figure 7). Approximately 80% of students using the article choice tutor and 75% of students using the explanation choice tutor requested hints on fewer than five problems (out of a total of 56 problems) and approximately 50% from each group never requested a hint. This suggests that any similiarities seen in learning between the two conditions is not a result of students seeing similar hints.

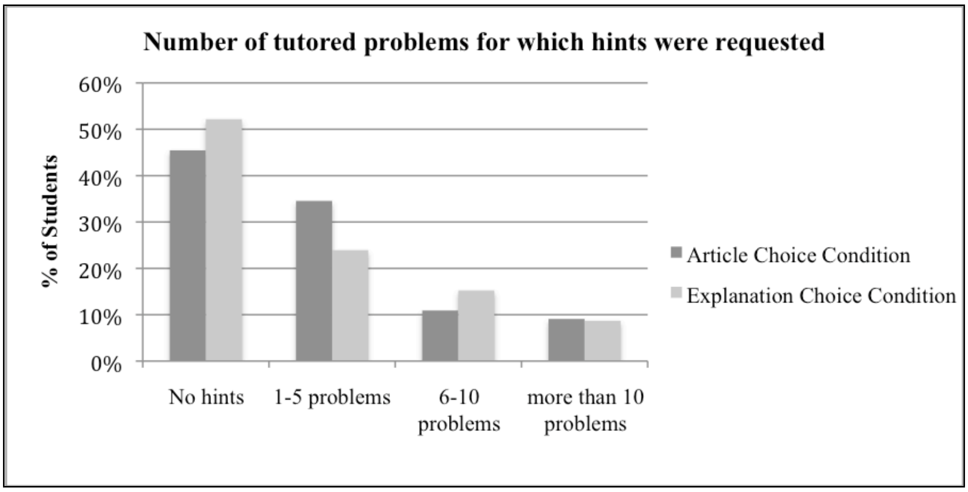



Figure 7: Overall, students using either tutoring system rarely asked for a hint to complete the problems.

## Effects of First Language

Finally, in order to determine if a student's first language (L1) had an effect on learning or if there was an interaction between tutoring condition and first language, we first classified participants' first languages into one of three categories: L1 has both a definite and indefinite article, L1 does not have a definite but does have an indefinite article, or L1 has neither a definite nor indefiniet article (Dryer, 2008) and then repeated the MANCOVA analysis described above, adding first language category as an additional independent variable. Results show no main effect for first langauge ($F(4, 104) = 1.31$, $p = 0.270$) and no interaction between first language and tutoring condition ($F(12, 104) = 0.683$, $p = 0.765$).

# Discussion

Both tutoring conditions were successful at increasing students' procedural knowledge (article choice) and declarative knowledge (explanation choice). These results suggest that students can learn from self-explaining not only in math and science domains, but also in new domains like second language grammar learning. The self-explanation prompts help students both to correctly use articles and to explicitly identify the rules driving these decisions. Further supporting this finding is the timing data that shows students in the explanation choice condition are solving the posttest problems just as quickly as their article choice counterparts. Thus, self-explanation improves both facets of fluency, speed and accuracy, just as well as practice.

While one might expect students in the article choice condition to improve on the article choice posttest and students in the explanation choice condition to improve on the explanation choice posttest, the fact that we see cross-task transfer for both tutoring conditions is somewhat surprising, although not unprecedented. In their study looking at the effects of self-explanation in geometry learning, Aleven & Koedinger (2002) saw similar results in that students in both the self-explanation condition and problem-solving condition showed significant learning in their ability to problem solve *and* in their ability to provide explanations or reasons. One key difference is that in the geometry study, students in the self-explanation condition had practice with both procedural (problem solving) *and* declarative (explaining) skills, while in our study, students in the self-explanation condition did not have any procedural (article choice) practice and were tutored on declarative skills (explanation choice) only. We hypothesize that the combination of worked examples (i.e. seeing the correct article highlighted in the problem sentence) and self-explanation prompts encourages students to focus on the features of the sentence that are important for making article decisions. They are then able to recall and apply these rules when presented with procedural items on the posttest.

We see a similar but smaller effect for students in the procedural (article choice) condition in that despite having no tutored practice on choosing explanations, they show significant improvement on the declarative items in the posttest. We hypothesize that the procedural instruction may have led to an increase in declarative knowledge through students inductively learning the rules and perhaps a priming effect by the pretest. Since all students took the explanation choice pretest, which asked them to use declarative knowledge, students may have been looking for and inductively generating the rules from the tutored problems they completed, accounting for the small but significant increase in declarative knowledge.

One limitation of this work is the generalizability of the results. Since the entire study was conducted during one class period, we had limited time within which to collect data. Due to this limitation, less than 80% of the students in the explanation choice condition finished all five stages of the study. While we used propensity score matching to account for the attrition in the analysis, we cannot make strong claims about those students for whom we have no posttest measures. As noted by the classroom teachers, students who did not finish likely fell into one of two categories: those with low prior knowledge and those with high prior knowledge who were very diligent. Perhaps self-explanation is not appropriate for all students. The metalinguistic challenges involved in choosing explanations in a foreign language might be too difficult for the low prior knowledge students, and asking very diligent students to self-explain each step may be an inefficient use of class time. To address these issues, on-going research is investigating interventions designed to foster feature-focusing behavior while minimizing metalinguistic difficulties as well as interweaving procedural and declarative tasks in order to reduce the amount of overall time needed to complete the tutor.

In conclusion, this study is one of the first to investigate the effects of self-explanation on language learning and presents promising results that prompted self-explanation helps students with procedural skills and enables them to develop explicit knowledge of the rules. Further work is needed to investigate the robustness of this effect and the extent to which this knowledge transfers to real-world production.

## References

Aleven, V. & Koedinger K. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, *26*, 147-179.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M.T.H., de Leeuw, N., Chiu, M.H., LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439-477.

Cole, T. (2000). *The Article Book: Practice Towards Mastering a, an, and the*. University of Michigan Press.

Conati, C., & VanLehn, K. (2000). Further results from the evaluation of an Intelligent Computer Tutor to Coach Self-Explanation. Proceedings of the 9th International Conference on Intelligent Tutoring Systems, 304-313.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and Second Language Instruction*, 206-257.

Dryer, M. (2008) Definite Articles. In: In: Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard (eds.) The World Atlas of Language Structures Online. Munich: Max Planck Digital Library, chapter 37. Available online at http://wals.info/feature/37. Accessed on October 28, 2009.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford University Press.

Hake, R. (1998). Interactive-engagement versus traditional methods: a six-thousand- student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66, 64 – 74.

Koedinger, K., Aleven, V., Heffernan. T., McLaren, B. & Hockenberry, M. (2004). Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration. Proceedings of 7th Annual Intelligent Tutoring Systems Conference. Maceio, Brazil.

Master, P. (1997) The English Article System: Acquisition, Function, and Pedagogy. *System*, 25(2), 215-232.

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, 38, 1-30.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis*. Language Learning*, *50,* 417-528.

Renkl, A. (1999) Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology in Education, 16*, 477-488.

Rosenbaum, P. & Rubin, D. (1984) Reducing the bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79, 516-524.

Roy, M. & Chi, M.T.H. (2005). Self-explanation in a multi-media context. In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning*, 271-286.

Williams, J. & Evans, J. (1998). What kind of foucs and on which forms? In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition*, 139-155.

## Acknowledgments

Special thanks to Howard Seltman for help with the analyses, the PSLC CTAT and DataShop teams and our partner teachers. This work was supported in part by the PSLC (National Science Foundation Award SBE-0836012) and by the Institute of Education Sciences, U.S. Department of Education (Grant R305B040063).

# Stressed yet Motivated: Web-Based Peer Assessed Competition as an Instructional Approach in Higher Education

**Abstract:** The study explores peer assessed artifact competitions conducted in a college setting. The competitions utilized a web-based environment for designing and enacting collaborative online activities. Since peer assessed team competitions are extremely effective in promoting learning yet are rarely conducted in higher education the study aimed at disclosing students' feelings about such potentially stressful circumstances. The peer assessed competitions were favorably accepted by most of the students who claimed that it improved their performance. Moreover, students level of stress was correlated with putting more efforts into the products and feeling that as a result its' quality improved. In addition, peer assessments were highly correlated with the instructor's assessments and students reported they trusted them regardless the rating received. Finally, Fear of Failure, as a personality trait, predicted which students would report being anxious about the competition. Implications for supporting such students are discussed.

## Introduction

Research indicates that team competition as an instructional approach produces higher learning performances than cooperative learning or competition among individuals (Slavin, 1980; Slavin et al., 1984; Fu et al., 2009; Ke & Grabowski, 2007). In team competitions a certain number of individuals cooperate as a group to compete against other groups. Such competitions are reported to be designed and successfully applied across various higher education programs such as Business Management (Corner et al. 2006; Casile, & Wheeler, 2005), Engineering (Cramer & Curten, 2005; Sansalone, 1990), Computing (Fu et al. 2009), and Instructional Design (Kinzie et al. 1998; Rowland, 1994). For instance, Corner et al. (2006) describe a case team competition that takes place every year where business management students analyze real-world complex case faculty members produced with the help of local businesses. Students give excellent evaluations to this activity and faculty members believe it is a wonderful experience. Cramer & Kurten (2005) too, describe a team competition occurring each year - where engineering students design , develop and test a Canoe made of concrete - as a peak experience in the program. However, all these studies ignore one of the most important pedagogical resources relevant to team competition – peer assessments; typically, students' teams are assessed and ranked by juries or faculty member, but not by the students themselves.

In higher education, peers have almost no input regarding their friends' projects flaws and merits, ways to improve it, etc. Yet, socio-constructivists theories argue that peers are one of the most influential factors in knowledge construction (Cole, 1996; Lave & Wenger, 1991). Peer assessment, in particular, has been shown to have many advantages. Studies have indicated that peer assessment assists students to create higher quality artifacts, as a consequence of better understanding of assessment criteria which they use when they play the role of assessors (Falchikov, 2003; Smith et al. 2002; Kali & Ronen, 2008). Since peer assessments are rarely used in competitive settings in higher education there is hardly any study that explored how students feel about exposing their work to peers under such circumstances. One of the purposes of this study is to answer this question.

Assuming that peer assessed competitions might have pedagogical advantages, orchestrating it (assigning artifacts for peer assessments, calculating and publishing the results, etc.) is logistically complex, hence, decrease faculties' incentive to embrace such an initiative. Recent e-leaning environments may provides an efficient solution to this challenge, by facilitating the design, orchestration and enactment of peer assessed competition activities, especially when dealing with artifacts that are produced in digital formats.

The CeLS (Collaborative e-Learning Structures), a web-based environment for designing and enacting collaborative online activities (Ronen et. al., 2006), provides an efficient solution to such a challenge. Using a friendly and intuitive interface lecturers use the CeLS to plan the competition and afterwards the successive stages are automatically handled by the system, e.g., students submit their product (the environment is designed to absorb multimedia artifacts), then, the environment randomly assigns a predetermined number of anonymous artifacts for each student to assess, then, the students submit their assessments. Finally, each artifact is publicly presented with peers' mean score plus anonymous verbal comments, adjacent to it. Typically each student receives scores and verbal comments form about 10-20 peers - depending on number of students in the class and number of artifacts each student is required to assess. Such a process allows for each artifact to be analyzed from multiple perspectives and gives each student a rich and multifaceted feedback. Since the CeLS environment supports the lecturer in structuring the activity and automates its enactment - using peer assessed competition as an instructional strategy hardly puts extra demands on the lecturer. In our study the CeLS environment was used to handle peer assessed competition activities in college level courses.

## The study

This study explores the pedagogical aspects of a peer assessed competition as an instructional approach in higher education. Our aim was to assess the quality and perceived value of the feedbacks students give their peers, and to figure out some of the motivational factors related to the fact that students know that their products will be assessed and ranked by their peers, and eventually will be publicly discussed in front of the whole class. How do students feel about such circumstances? Do they feel intimidated? Does it energize them? Do they put more effort into the project then they would normally do? Are personality traits, such as fear of failure, related to the ways they deal with the knowledge their products will be 'exposed' and assessed by their peers? Does fear of failure inhibit competitiveness? Do they appreciate their peers' feedback? In particular the study addresses three issues:

- How reliable are peer assessments? Do they correlate with the instructor's assessments? Do students trust peer judges?
- To what extent students feel that the web-based competition motivated them and encouraged them to submit better artifacts or, on the contrary, inhibited them and harmed their performance?
- Is Fear of Failure, as a personality trait, related to students' attitudes toward web-based competitions? Better understanding of the psychological dynamics related to such competitions might help us support students who dislike such activities and experience excessive stress related to them.

### Participants & Activities

Participants were 1st and 2nd year undergraduate students in an Instructional Technology B.A. program in a technological college. The program's curriculum is fully dedicated to instructional technologies (there are no other majors or minors) and students are involved in many team projects during their 3 years study. In the first two years all students take the same compulsory courses, while elective courses are offered only in the 3rd year. As a result of these circumstances students' cohorts form very cohesive groups. This fact may have implications on the ways students' experienced different aspects of the activity, such as the need to assess peers, the experience of being assessed by peers, and the experience of overt competition, as will be demonstrated in students' quotes reported later on.

41 1st first year students participated in an Introduction to Psychology course. As part of the social psychology unit students were asked to apply attitude-change theories and design a poster and a brochure that encourage parents to consider sending their children to schools that adopt constructivists' principles. Peers' assessment and competition in this group dealt with the poster and its efficacy and potential impact for raising awareness and attitude-change.

44 2nd year students participated in a Web based Inquiry Learning course. As part of the course the students were asked to design and develop a prototype of a WebQuest (Dodge, 1995) dealing with an historical dilemma (related to world war II) the lecturer provided. Peers' assessment and competition in this group dealt with the WebQuest design. Students were encouraged to use a rubric (Dodge, 2001) to support the design and the assessments process. The rubric contained assessment dimensions such as the effectiveness of motivational strategies used to provide an engaging webquest, the appropriate use of collaborative learning strategies, the sophistication of the task design in terms of encouraging higher-order thinking, etc.

## Method & Tools

At the beginning of the semester students completed a Fear of Failure personality questionnaire (Performance Failure Appraisal Inventory, Conroy, Willow & Metzler, 2002). The questionnaire measures the strength of individuals' beliefs in five aversive consequences of failing: fear of experiencing shame and embarrassment, fear of devaluing one's self-esteem, fear of having an uncertain future, fear of important others losing interest and fear of upsetting important others. The scores obtained from these five consequences are moderately- to strongly-correlated with each other and their common variance can be modeled with a single higher-order factor representing a general fear of failure. The breakdown of Fear of Failure to such 5 scales enables one to better understand the exact cause of fear and properly address it (Conroy, Willow & Metzler, 2002). Fear of failure, general, is believed to intimidate need for achievement and compositeness (McClelland, 1961).

During the semester students participated in the competition activity and were asked to assess their peers' projects and grade the artifacts. The specific activities and their web-based implementation are detailed in the next section. The competition activities were an integral part of the courses; participation as assessors was credited while the actual grades for the artifacts were given *only by the instructor* and students' assessments and ranking did not influence it.

Following the experience students answered a questionnaire (14 multiple selection items and one open question) reflecting on their feelings and attitudes related to participating in the web-based competition and their preferences regarding such projects. The questionnaire asked students to provide Likert type ratings to items such as: "the fact that my artifact was rated by my peers: stressed me, made me put more effort into the project, made me conduct more improvement trials, ended up improving my artifact", "peers' assessments seem valid to me", "I would like to have more such competitive-like activities". The open question asked for students' opinion about the competitive activity.

## Web-Based Competition: Instructional Design and Implementation

The competition activities were performed with CeLS (Collaborative e-Learning Structures), a web-based environment for designing and enacting collaborative online activities (Ronen et. al., 2006). The system offers content free templates and a searchable repository of sample activities that were implemented with students. Teachers can explore these resources and adapt them to suit their needs or create new activities from basic building blocks. CeLS unique feature is the ability to design activities that use learners' artifacts from previous stages according to various Social Settings (e.g., one team accesses another team's project, a whole class gives improvement ideas to one team, etc.). The Social Settings determine which and how many artifacts would be presented to each participant for further interaction. This feature is exploited in order to design the competition activities and to facilitate their enactment in a real setting. The general structure of a competition activity consists of three stages (Figure 1):

*Stage 1* presents the activity and provides a dedicated place for the artifacts submission. The submission interface is adapted by the teacher to suit the artifact's requirements: in the posters competition, in our study, the interface invited students to summit JPG files up to 300kB and in the WebQuest competition a the interface provided a location for submitting a proper link to a website. The artifacts submitted can be individual or group products, as defined by the Social Settings. In our case the artifacts were group products. This definition would enable either of the group members to submit the artifact then ensure that in the assessment stage a student would not be presented with her own group's artifact.

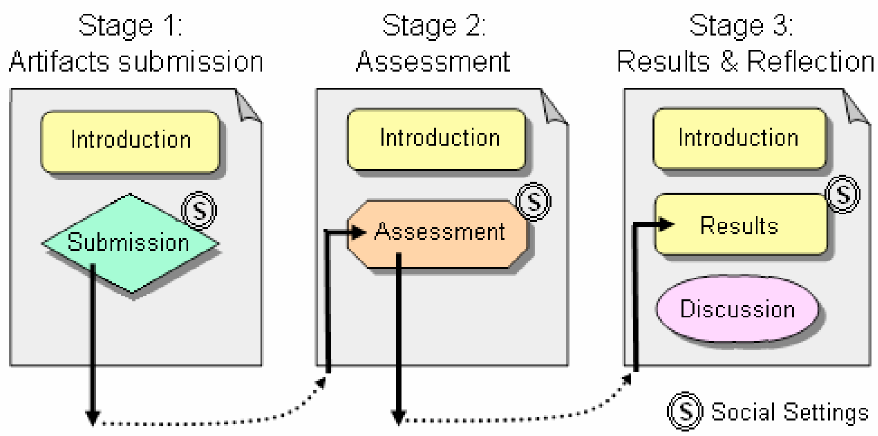



Figure 1. The general structure of a competition activity implemented with CeLS.

*Stage 2*: The Assessment stage starts with relevant instructions, then (some or all) peer artifacts are presented for voting (see Figure 2a). The artifacts are presented anonymously to the peers. If the activity involves assessing many artifacts (as in the Poster competition) or if the artifacts are complex and their assessment requires considerable effort (as in the WebQuest competition), it would be advisable to restrict the load and present each judge with a limited number of artifacts. Therefore, each judge in the Poster competition was asked to assess 8 artifacts (Figure 2a) while in the WebQuest competition only 5 artifacts were assessed by each student. Students were allowed a couple of days for these assessments. The Assessment interface in our competition activities was adjusted so that each assessor had to provide an overall grade and verbal justifications and explanations. If an activity would require a more detailed evaluation addressing various criteria, the Assessment interface could be a questionnaire or a rubric.

*Stage 3* presents the competition overall results (see example Figure 2b). The teacher can show the assessment details (grades and justifications) for all artifacts (presented anonymously) to all students or to present each participant only with the details for his own product.

Figure 2 presents partial sample screens from Stage 2 and Stage 3 of the poster competition.

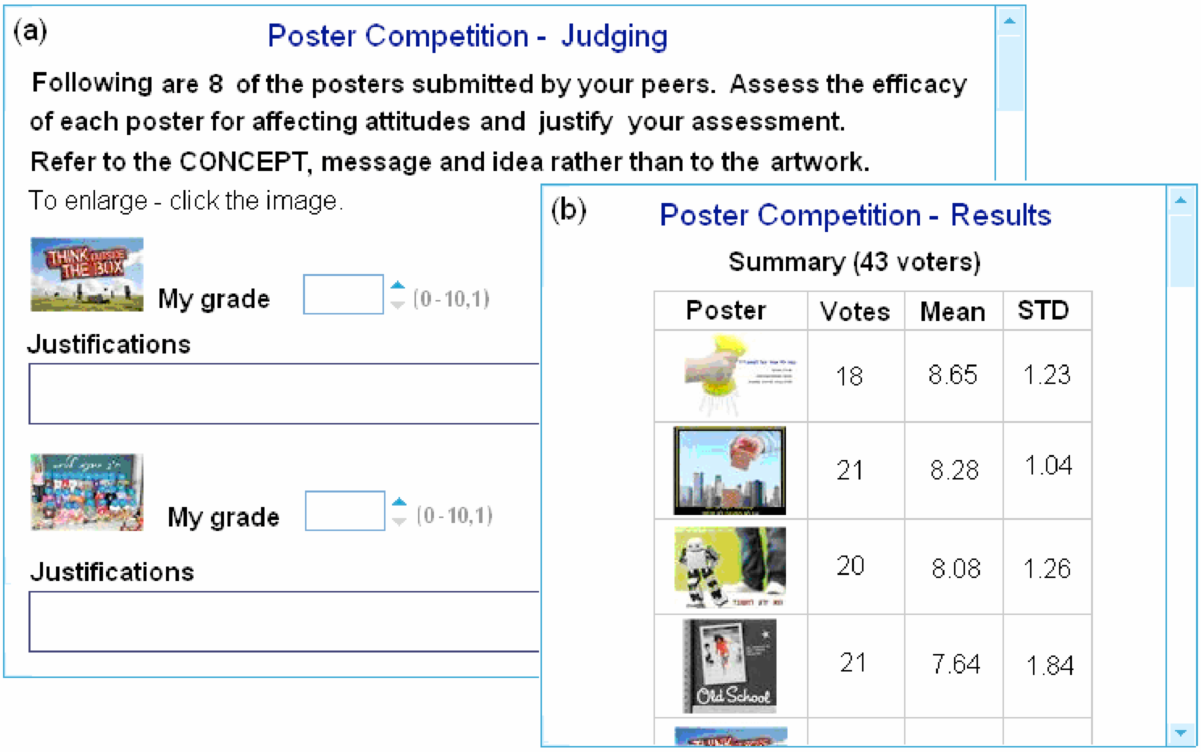



Figure 2. Sample screens from the poster competition. (a) Stage 2 (b) Stage 3.

## Results

First we will present data regarding the quality and worth trustfulness of peers' assessments and to what extent they feel their peers judged their artifact fairly. Then we will present students' attitudes toward competing and being assessed by peers and their perception of the impact of the competition approach on the quality of the artifacts. Finally, we will report about the relationships between Fear of Failure as a personality trait and students' attitudes toward web-based competition activities.

### Students as Assessors

Table 1 presents the correlations between students' mean ratings of artifacts and the lecturer's independent ratings. Similar to Kali & Ronen (2008) findings, students' mean ratings seem to be valid and correlate nicely with the instructor's grades.

Table 1: Correlations between students' mean ratings of artifacts and the lecturer's independent ratings.

| Activity | # of Artifacts | r | p |
|---|---|---|---|
| Year 1: Poster Competition | 20 | 0.82 | 0.000004 |
| Year 2: WebQuest Competition | 24 | 0.68 | 0.0001 |

There is a noticeable difference between the two correlations presented in Table 1. The higher correlation between students and instructor in the Poster competition is probably due to the fact that the artifact is less complex and less multi dimensional then the WebQuest one. In the WebQuest activity the assessment process was much more demanding. As a result some students were biased by salient features (such as interface and visual design) rather than assessing factors such as the pedagogical value of the WebQuest, resulting in a lower correlation between students' and instructor's assessments.

Students' comments to artifacts were interesting, non-repetitive, and provided an enriching and insightful analysis of the artifacts. It seems that a feedback given only by the instructor could not provide the intellectual and emotional impact of such a multiple perspectives feedback to many artifacts (see examples in Figure 3).

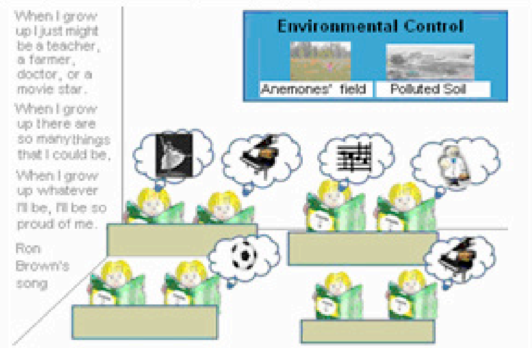
| A poster ranked highly by peers and instructor | | A poster ranked poorly by peers and instructor | |
|---|---|---|---|
|  | |  | |
| Peer grade | Justification | Peer grade | Justification |
| 9 | A very nice way to represent symbolically the idea (the sponge), plus a strong appeal to emotion. The verbal information is also very focused; it offers an alternative which emphasizes experiential, constructivist learning principles. Alfi Kuhn's ideas are very well represented in this poster. | 6 | It seems that the message intended to be delivered is that classroom materials are neither interesting nor relevant to children. They have big dreams and should be provided with means to reach these dreams. An alternative is missing: What should be done in order to bring relevance to the educational system? |
| 9 | I love the message which brings me back to the time where I studies History and had to learn by heart many 'dry' facts. It stresses the point that students absorb the material using rote learning and then they 'vomit' it. Loved it. | 5 | Overload of text and of visual elements as well. There is no clear saying. It is impossible to understand from the poster what is the problem and how to deal with it. The graphic design isn't impressive either. |
| 8 | This poster may cause actual attitude change! Parents would not want their children to be like a sponge with no understanding! Parents would like their child to learn, develop, understanding. The idea is very well capsulated and is very clear. The message will have a strong emotional impact on parents. | 5 | Although it is obvious that a lot of work and thought was put in the poster no real message is delivered, particularly when the topic we are dealing with is considered. One has to 'shock' the audience more heavily with clear-cut sentences that indicate what is the problem and what should be done to fix it. |
| 9 | Uses humor in an effective strategy. Does not use a simplistic one-sided attitude change strategy but rather a sophisticated message delivered lightly and clearly. | 7 | Good idea but it is somewhat unclear – the conformity on the one hand versus the over-emphasized wish to be unique on the other – but who/what could help children become unique? What is the next step? |




Figure 3. Examples of posters and few sample peers' assessments.

Most students in both classes (about 80%) felt that peer assessments to their own artifacts were valid and fair. No significant correlation was found between the appreciation of peer grades' validity and the actual grades granted by the peers, meaning that this view was shared also by students whose artifacts were judged less favorably by their peers.

## Competition as an Instructional Strategy – Student's Views

The analysis of 1st and 2nd year students' responses to the reflective questionnaire and interviews revealed very similar distributions, therefore we shall refer to both groups as a whole (N=85).

Most students felt that the awareness to the fact that their artifacts would be exposed and assessed by peers resulted in a better product (Table 2).

Table 2 : Perceived impact of the awareness that the artifacts would be assessed by peers on artifacts' quality.

| Perceived impact on artifacts' quality (%) | | | | |
|---|---|---|---|---|
| None | Little | Some | Large | Very large |
| 14 | 20 | 40 | 22 | 4 |

Only few students (4%) reported that the competition was somewhat "paralyzing" resulting in a detrimental effect upon their artifacts.

As could be expected, self reported motivation caused by the competition (items such as "the fact that my artifact was rated by my peers: made me put more effort into the project") was significantly correlated with the perception of positive effect on products' quality (r=0.72 p<0.000001). The more interesting finding is related to the perception of stress caused by the competition activity (based on the item: "the fact that my artifact was graded by my peers stressed me"). Even if no correlation was found between the degree of stress caused by the activity and the perceived impact on products' quality, the estimation of stress was significantly correlated with the perception of effort devoted to the creation of the artifacts (r=0.33 p<0.001), meaning that students who admitted to being more stressed also felt that they have devoted more time and effort to the activity. The beneficial effect of "some stress" is demonstrated in Table 3 that presents the student's perceptions of the competition activity as motivating and as stressing.

Table 3: Student's perceptions of the competition as a motivating and as a stressing activity (N=85).

| | | Motivation (%) | | | |
|---|---|---|---|---|---|
| | | none | some | much | Total |
| Stress (%) | none | **4** | **23** | **23** | 50 |
| | some | **2** | **20** | **24** | 46 |
| | much | **1** | | **3** | 4 |
| | Total | 7 | 43 | 50 | 100 |

Even if most (60%) would favor using competition activities in academic courses, in their detailed comments many recommended not to "overuse" this strategy and to restrict it to once in a semester in order not to impose an "exaggerated" workload. The quantitative aspects were supported by students' written comments. Following are few examples of students' opinions regarding peer judged competition as an instructional method:

**Positive opinions:**

- It is very motivating. It made me put lots of effort. I think that without the competitive factor I wouldn't have achieved such a success in this WebQuest activity. In addition the CeLS environment allowed us all to see each others' products, which is very nice for comparing and learning.
- The number of WebQuests we had to assess (5) was just right. I would not ask students to assess more then that. I believe competition is always a good idea. The fact that the leading products were presented and discussed in class made me feel really good after all the effort I have put in it.
- I am a competitive type so this method really helped me to achieve a meaningful product. I was enriched by viewing the elaborated products of my peers. To summarize, it was fun and educative.

**Ambivalent opinions:**

- I think the competitive activity is useful, efficient and contributing as long as a rubric is provided and the grading process is led by it. In this WebQuest activity we were encouraged to use a rubric. According to the competition results, it seems that not everyone used it - so there isn't enough uniformity in the grades given by peers and in some cases there weren't sufficient explanations to support assigned grades.
- I think that to some extent the competition did stimulate interest and encouraged teams to produce better WebQuests – yet, personally I believe that it is not always good to conduct such competitions since it might create tensions between class members, create uncomfortable situations, or unreliable results.
- I don't like so much competitive activities. Working in a primary school I saw cases where students gave up in advance, since they thought they don't have a chance. On the other hand – I think that from time to time, such an experience could be a fun and refreshing.
- Competition could paralyze people with low self-confidence – yet, with right team work such a problem might be less dominant. I believe in competition, believe it motivates, and produces higher quality products. Of course one shouldn't exaggerate and put people into too much pressure.
- Competition is a good yet problematic method. It was difficult for me to assess close friends from our class. In addition, competitive students would not want to give high grade to others that might surpass them. In a small and cohesive class as ours - it is hard to critique others.

**Negative opinions:**

- Some people need to learn how to provide feedback to others – and the lecturer needs to stress it.
- After reading the feedback given to our WebQuest I was pissed off. I realized that people drastically punished us for criteria that weren't relevant. It seems that people didn't realize what the purpose of the activity was. Some groups created a fully functioning web site rather then a prototype. As a result feedback focused on usability and visual design aspects rather then the criteria defined by the rubric provided.
- The problem with competition is that it puts pressure on those who are highly anxious. I am sorry that in our class (1$^{st}$ year, poster project) anxiety is more salient then healthy competition. But maybe it is only our class and when time passes it will change.

One can see the competitive activity raised some intense emotions. Some students loved it while others were skeptical about their peers' feedback and worried about augmenting the anxiety level in class. It seems that most of the students felt that activity was refreshing and fruitful yet shouldn't be implemented too often.

## Fear of failure and students' attitudes toward Web-based competition

Fear of failure was found to be related to participants' negative attitudes towards web-based competition. There was a considerable correlation ($r=.47$ $p<.01$) between Fear of Failure (sub-scale related to devaluing one's self-esteem) and participants' reporting that the competition paralyzed them. Similarly, Fear of failure (sub-scale related fear of upsetting important others) was correlated to students' preference to have peers assessments without numeric grades and competitive aspects ($r=.4$ $p<.01$). On the other hand, Fear of Failure (sub-scale related to fear of experiencing shame and embarrassment) was correlated with reporting about conducting more improvement trials ($r=.32$ $p<.01$).

It seems, thus, that fear of failure has different dimensions and is not a holistic psychological phenomenon with uniform consequences. Some dimensions (e.g., devaluing one's self esteem, or fear of upsetting important others) might inhibit students and consequently cause a decline in performance, whereas others (such as fear of experiencing shame and embarrassment), on the contrary, might boost students' motivation and make them invest more effort in their artifacts. Such fine differentiations might give instructors a clue regarding ways to support students who experience excessive stress related to peer-assessed competitions. For instance, instructors could encourage anxious students to try and diminish the direct, and at times painful, connection between their products and their self-esteem. Alternatively, an instructor could gently suggest that as grownups, important others might perceive such students more favorably then dominant persons related to their past.

## Discussion

The findings support previous reports which provide evidence that peer assessments are trust worthy, reliable, and in many cases highly correlate with instructors' assessments (Kali & Ronen, 2008). Students' verbal comments in our study demonstrate the emotional and intellectual advantage of feedback provided from multiple perspectives as compared to feedback provided exclusively from the instructor. Students' comment to peers' artifacts seem to be enriching, interesting, stimulating and honest. Despite the concerns of few, most students reported they trusted their friends' feedback, and this was true also for students who did not do so well.

Most students believe that the awareness to the fact that their products would be assessed by their peers boosted their motivation and as a result they submitted better artifacts. The fact that the experience of stress was positively correlated with effort invested supports the famous inverted U theory (e.g., Muse et al., 2003) which claims that moderate amounts of stress improves performance. However, our data suggests that some students might experience intense stress related to peer ranking and competition; Students who tend to devalue themselves as a result of failure, and students who tend to fear they upset important others when they fail – seem to have experienced the competition as paralyzing and would prefer not be ranked by their peers. On the other hand students who tend to feel embarrassed as a result of failure seem to put more effort and improve their performance. These findings shed some light on the possible dynamics causing some students to be paralyzed by competitive-like activities and provide some clues on how an instructor might support them.

The academic grading system is based on competitive sorting. Many believe that competition is detrimental to learning and to intrinsic motivation and call for the minimization of its effects by means of educational strategies such as collaborative learning (Kohn, 1992). Do competitive learning activities augment the harmful effects of grading and competitive sorting? Our quantitative and qualitative data supports studies that claim the team competition is and effective and engaging experience (Slavin, 1980; Slavin et al., 1984; Fu et al., 2009; Ke & Grabowski, 2007). Recent motivation theories (Covington & Wiedenhaupt, 1997) argue that intrinsic and extrinsic motivations are two independent dimensions (they suggest a quadripolar model rather than a bipolar one). Thus, despite the fact that competition factors might increase extrinsic motivation, one might still be intrinsically involved in a task, as long as it is intriguing, fun, and more similar to 'play' rather then 'work' (Covington & Wiedenhaupt, 1997). One can be stressed yet intrinsically motivated at the same time. Nemerow (1996) reached similar conclusions: In a study applying both competitive and noncompetitive games, students were surveyed to find out how they felt about the games and what they learned from them. Results indicated that competitive games helped students improve self-esteem, peer relationships, and learning, yet, the students described the competition as motivating but also producing pressure.

The competitive learning experience in our study seemed to endorse a playful climate and to intrinsically engage most of our students and caused many of them to believe they improved their performance. Other higher education studies which experimented with similar, fun, team competition activities support this notion (e.g., Corner et al. 2006; Casile, & Wheeler, 2005; Cramer & Curten, 2005; Kinzie et al. 1998).

## Summary & Concluding Remarks

The fact that the team competition activities energized students in our study and made them believe that it improved their products might indicate that when a favorable social climate is created most learners might enjoy and benefit from the 'public' exposure and competitive situation involved with peer ranking and assessments. The rich and divers nature of multiple perspective feedback seems to have a good potential of augmenting learning processes and meta-cognitive self-assessment abilities (White & Frederiksen, 2000). Teachers in higher education should consider using more often such strategies in any discipline where the creation of original artifacts is relevant. As long as they succeed to create a playful, psychologically safe learning environment, chances are most students will get intrinsically involved, enjoy the experience, and as a result of assessing others' work, become more reflective about their own learning. Using a web-based environment such as CeLS takes care of the logistic hassles and increases the chances that faculty members will enjoy the experience as well.

## References


Cramer, S., Kurten, J. (2005). Inspire Future Engineers with the Concrete Canoe Competition. *Tech Directions, 65* (4), 17-19.

Casile, M. & Wheeler, J. V. (2005). The Magnetic Sentences Industry Game: A Competitive In-Class Experience of Business-Level Strategy. *Journal of Management Education, 29*(5), 696-713.

Cole, M. (1996). *Culture in mind.* Cambridge, MA: Harvard University Press.

Conroy D. E., Willow J. P., Metzler J. N. (2002). Multidimensional Fear of Failure Measurement: The Performance Failure Appraisal Inventory. *Journal of Applied Sport Psychology, 14* (2), 76-90.

Corner, P. D., Bowden, S., Clark, D., Collins, E., Gibb, J., Kearins, K., Pavlovich, K. (2006). Grounded Learning from a Strategy Case Competition. *Journal of Management Education, 30*(3), 431-454.

Covington, M. V., & Wiedenhaupt, S. (1997). Turning work into play: The nature and nurturing of intrinsic task engagement. In R. P. Perry & J. C. Smart, (Eds.), *Effective teaching in higher education: Research and practice*, (pp. 101-114). Agathon Press, New York.

Dodge, B. (1995). WebQuests: A technique for Internet-based learning. *Distance Educator, 1*(2), 10-13.

Dodge, B. (2001). *A Rubric for Evaluating WebQuests.* Retrieved October 12, 2009 from http://webquest.sdsu.edu/webquestrubric.html

Falchikov, N. (2003). Involving students in assessment. *Psychology Learning and Teaching, 3*(2), 102-108.

Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A meta-analysis comparing peer and teacher marks. *Review of Educational Researc, 70*(3), 287-322.

Fu, F.L., Wu, Y.L., Ho, H.C. (2009). An Investigation of CoOpetitive Pedagogic Design for Knowledge Creation in Web-Based Learning. *Computers & Education, 53*( 3), 550-562.

Kali, Y., & Ronen M. (2008). Assessing the assessors: Added value in web-based multi-cycle peer assessment in higher education, *Research and Practice in Technology Enhanced Learning, 3*(1), 3-32.

Ke, F., & Grabowski, B. (2007). Game playing for Math learning: Cooperative or not? *British Journal of Educational Technology, 38*(2), 249–259.

Kinzie, M.B., Hrabe, M.E., Larsen, V. A. (1998). An Instructional Design Case Event: Exploring Issues in Professional Practice. *Educational Technology Research and Development*, *46*(1), 53-71.

Kohn, A. (1992). Resistance to Cooperative Learning: Making Sense of Its Deletion and Dilution. *Journal of Education, (174)*2, 38-56.

Lave, J. & Wenger, E. (1991). *Situated learning. Legitimate peripheral participation.* Cambridge: Cambridge University Press.

McClelland, D. C. (1961). Methods of Measuring Human Motivation, in J.W. Atkinson, (ed.), *The Achieving Society*, pp. 41-43. Princeton, N.J.

Muse, L. A., Harris, S. G., & Feild, S. H. (2003). Has the Inverted-U Theory of Stress and Job Performance Had a Fair Test*? Human Performance, 16*(4), 349–364.

Nemerow, L.G. (1996). Do Classroom Games Improve Motivation and Learning? *Teaching and Change, 3* (4), 356-566.

Ronen, M., Kohen-Vacs, D. & Raz-Fogel, N. (2006). Adopt & Adapt: Structuring, Sharing and Reusing Asynchronous Collaborative Pedagogy. In Barab, S., Hay, K., & Hickey., D. (Eds), *Proceedings of the 7th international conference on Learning sciences* (pp. 599-606). International Society of the Learning Sciences (ISLS).

Rowland, G. (1994). Educating Instructional Designers: Different Methods for Different Outcomes. *Educational Technology, 34*(6), 5-11.

Sansalone, M. (1990). Teaching Structural Concepts through Case Studies and Competitions. *Engineering Education, 80*(4), 472-475.

Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research, 50(2),* 315–342.

Slavin, R. E., Leavey, M. B., & Madden, N. A. (1984). Combining cooperative learning and individualized instruction: Effects on student mathematics achievement, attitudes, and behaviors. *The Elementary School Journal, 84*(4), 408–422.

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International, 39*(1), 71-81.

White, B. Y., & Frederiksen, J. R. (2000). Technological tools and instructional approaches for making scientific inquiry accessible to all. In M. J. Jacobson & R. B. Kozma (Eds.), *Innovation in science and mathematics education* (pp. 321-35). Mahwah, NJ: Lawrence Erlbaum Associates.

# Assessing Change in Learner's Causal Understanding Using Sequential Analysis and Causal Maps

Allan Jeong, Florida State University, Department of Educational Psychology & Learning Systems, Instructional Systems Program, Stone Building Rm. 3205E, ajeong@fsu.edu

**Abstract**: New methods are needed to assess how group discourse triggers changes in causal maps and to measure and visualize across time how changes in causal maps of individuals or the collective group progress toward the group or target map. The software tool, jMAP, was developed to enable learners to individually produce causal maps, download and aggregate the maps across learners. It aggregates learners' maps to reveal similarities between individual/group maps, the percentage of maps sharing particular causal links, average causal strength assigned to specific links, and degree of match between maps of the collective group and target/expert diagram. jMAP produces data to create transitional state diagrams for visualizing how causal maps change over time and the effects of specific dialogic processes. This paper presents findings from two case studies to illustrate how jMAP can be used to support the assessment of causal understanding, and identifies areas for future research.

## Introduction

Each one of us holds different beliefs and theories about the world. Learners' theories can be conceived, articulated, and assessed more efficiently in the form of causal maps - networks of events (nodes) and causal relationships (links) between events – than in the form of linearly written text. Some causal maps may be more accurate than others—depending on the presence and/or absence of supporting evidence; and some maps and the causal links within the maps may be more or less firmly held—depending on both the strength of the supporting evidence and the strength of specific causal relationships. Furthermore, causal maps are not fixed and unchanging. Instead, they are incomplete and constantly evolving; may contain errors, misconceptions, and contradictions; may provide simplified explanations of complex phenomena; and may often contain implicit measures of uncertainty about their validity (Seel, 2003). As a result, causal maps can change, but usually not randomly. That is, we presume that events trigger and provide the impetus for change. Causal maps and other similar forms of visual representations are being increasingly used to help assess learners' understanding of complex domains and/or learners' progress towards increased understanding (Spector & Koszalka, 2004; Nesbit & Adesope, 2006). However, the methods and software tools to measure how learner's maps change over time (Ifenthaler & Seel, 2005; Doyle & Radzicki, 2007) and how specific events (e.g., pedagogical discourse) trigger changes in learners' causal maps (Shute, Jeong, & Zapata-Rivera, in press) have not yet been adequately addressed.

To address some of these methodological challenges, Ifenthaler and Seel (2005) used transitional probabilities to determine how likely learner's maps (when examined at as a whole) changed in structural similarity across eight different time periods. Raters were given a specially designed questionnaire to determine if a learner's map at one point differed in structure from the learner's map produced from the most previous point in time. The study found that maps were most likely to change in structure at the early stages of the map construction process with the likelihood of changes dropping from one version to the next. However, Ifenthaler (2008) found that changes in scores on seven of nine measures of structural quality (e.g., total number of links, level of connectedness, average number of incoming and outgoing vertices per node) had no correlation to the degree to which the learners' maps matched the expert map. Not surprisingly, the one aspect of the learners' maps that did correlate to learning was the number of links shared between the learner's map and the expert map. These findings altogether suggest that measures used to gauge changes at the global level (where the unit of analysis is the map as a whole) and measures that are not scored in relation to a target map (e.g., expert or collective group map) may have little or no value as an assessment tool.

One alternative approach is to measure changes at a more micro-level using the node-link-node as the unit of analysis and unit of comparison between learners' and target maps. At this level, we can examine how likely links between specific nodes are to change from one state to another (e.g., strong vs. moderate vs. weak vs. no causal *impact*; or high vs. moderate vs. low *probability/confidence*) as maps change over time. We also see to what extent the observed changes in the values of each causal link converge towards the target causal link values present in the target map. For example, we expect that the causal link values for links representing learner's misconceptions (e.g., erroneous links *not* observed in the target map) or learners' shallow understandings (e.g., links between two nodes not directly related and/or better explained by inserting a mediating node) will converge towards a value of 0 (no causal link) over time, following a close examination and critical discussion of the causal relationships. At the same time, the expectation is that the causal link values of the links *not* observed in a learner's map (but present in the target map) will progress from a value of 0 to the value observed in the target map. Using the node-link-node as the unit of analysis enables a precise examination of how and to what extent observed changes in targeted links help and/or inhibit learners from achieving the target learning outcomes (e.g., deeper and more accurate understanding). Furthermore, this approach enables us to examine how specific interventions/events (e.g., depth of argumentation, the production of supporting evidence) affect the direction and magnitude of changes across links that are missing or links that are valid or invalid.

For example, Shute, Jeong & Zapata-Rivera (in press) conducted a study to examine the processes of collaborative theory construction in an online graduate course on instructional technology. In this study, each student used jMAP to individually construct a causal diagram (at the beginning, middle, *and* end of the semester) that explains the complex events and conditions (including intermediate events and their causal relationships) that determine when the use of media technology increases learning and achievement. In the causal maps, students assigned a strength value to each causal link (1 = weak, 2 = moderate, 3 = strong impact) based on personal experiences and empirical literature examined in the course. In addition, students were instructed to specify with each causal link added to the causal diagram the quality of evidence they possessed and/or compiled to justify the plausibility of each given link (0 = no evidence, 1 = weak evidence, 2 = moderate evidence, 3 = strong evidence). In this case study, the experimenters coded all the maps by hand and causal links into adjacency matrices because the early versions of jMAP did not perform this function. Once coded, jMAP was used to tabulate the sequential changes in causal links observed in each student's causal diagrams produced *prior* to and *subsequent* to collaborative work (identifying factors; collecting, annotating, and sharing supporting evidence; cross-examining the evidence; interpreting the evidence; consensus making).

The Discussion Analysis Tool (Jeong, 2005) was then used to sequentially analyze the data to produce the transitional state diagram presented below. The state diagram in Figure 1, for example, shows that 50% of all causal links that were assigned a strength value of one remained the same between the first and second, and between the second and third causal diagrams, when *no* evidence was presented nor discussed in the online group discussions to establish the plausibility and the strength of the link. In contrast, the state diagram on the right shows that when evidence was presented with the causal links, these same links with strength value of one were much more likely to remain the same (86% instead of 50%). Overall, this preliminary study illustrates how the jMAP environment – when combined with sequential analysis – can produce a potentially powerful method to studying the *processes* of theory construction and the factors and conditions that both support and inhibit the process.

This presentation will demonstrate the software tool called jMAP that can be used to identify differences between learners' causal maps, initiate collaborative argumentation to produce justifications for proposed causal links, and produces changes in learners' causal maps that better reflect/represent complex phenomena. Similar to the Cognizer program (Nakayama and Liao, 2005), jMAP enable each learner to produce a causal map with numerically weighted links (Figure 2) thus reducing unwanted biases and influences of other learners (Doyle & Radzicki, 2007). jMAP can then be used to: 1) automatically code diagrams into adjacency matrices; 2) upload, download, and aggregate multiple matrices so that the maps of an individual, the collective group, and/or expert (in any paired combination) can be graphically superimposed to determine for example the degree to which a learner's diagram matches that of the expert or the collective group at a given point in time; 3) view the superimposed maps across different time periods to both visualize and quantitatively assess how learners' causal diagrams change over time and the extent to which the changes are converging towards the maps of the expert or the collective group; and 4) sequentially analyze and visualize how particular learners' causal diagrams and causal understanding of complex phenomena change over time under different instructional processes, events, or conditions (Jeong, 2008).

jMAP also enable researchers and teachers to: (a) graphically superimpose an *individual* learner's map over the expert/target map to visually identify and highlight changes occurring over time in the causal maps of an individual or group of learners; (b) determine the extent to which the observed changes progress toward a target or collective model; (c) determine precisely where, when, and to what extent changes occur in the causal links within the causal maps; and most importantly; and (d) identify and measure how and to what extent specific events (e.g., viewing consensus data, discussing evidence, engaging in specific and critical discourse patterns) trigger changes in the causal links between various states (e.g., strong, moderate, weak, and no causal link) as demonstrated in Figure 3.

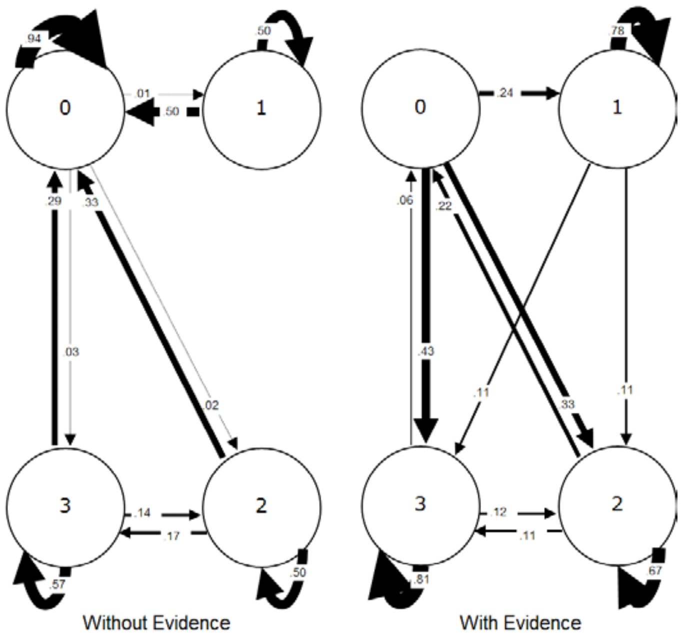



Figure 1. Transitional state diagrams revealing the direction and likelihood of changes in causal strengths when links are presented without vs. with supporting evidence

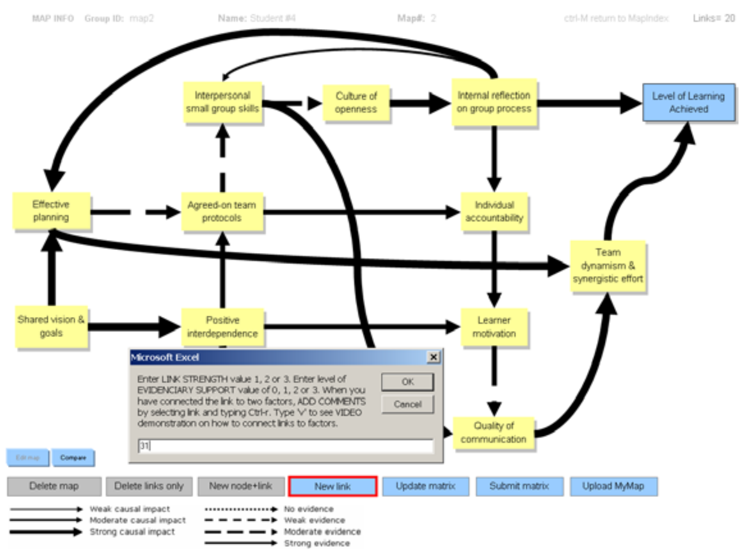
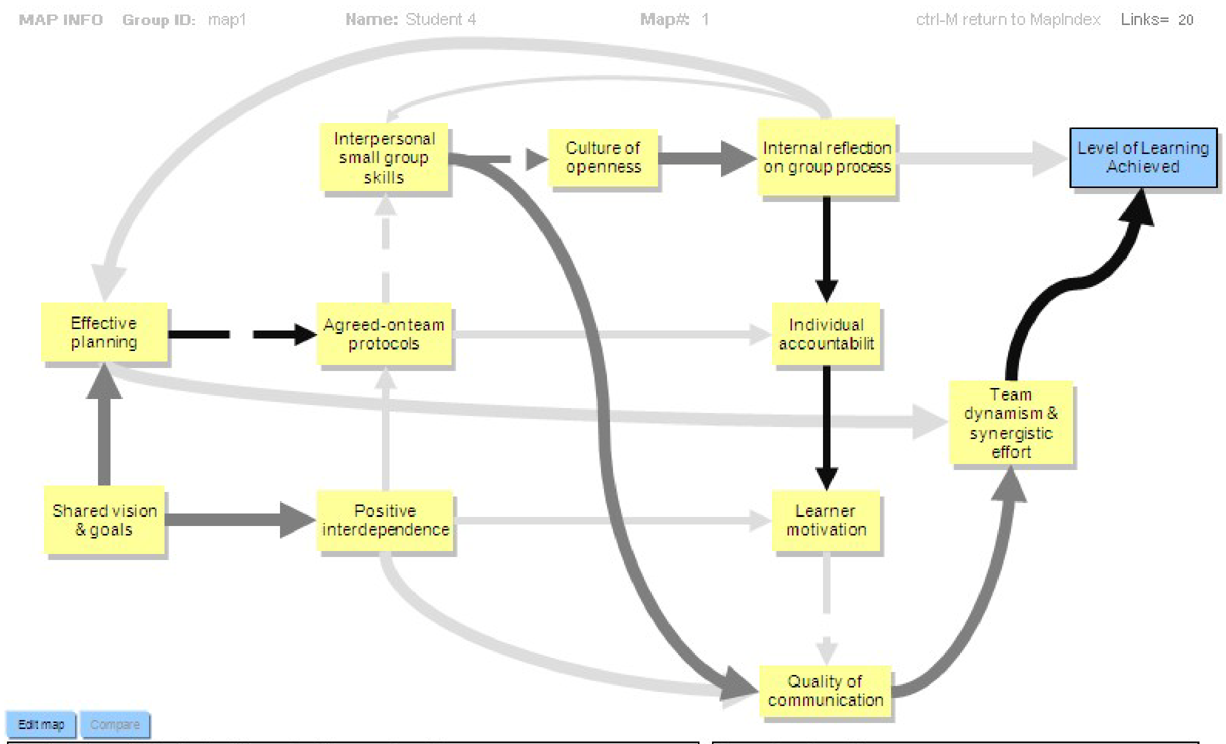



Figure 2. Drawing a causal map with jMAP using weighted links to specify strength of each causal relationship and dotted links to specify level of confidence or evidenciary support




| Navigational Tools for Visually Comparing Maps | | | |
|---|---|---|---|
| View my matrix | *ctrl-t* | | |
| View selected matrix | *ctrl-y* | *ctrl-u* | *ctrl-i* |
| Maps to view & compare with my map | g1 | 1-1 | 2-1 |
| *ctrl-g & ctrl-l scrolls thru MapIndex* | *ctrl-h* | *ctrl-j* | *ctrl-k* |
| Number of maps listed in MapIndex = 17 | | | |

| Quantitative Measures | |
|---|---|
| 4 | Links with strength values matching my map |
| 6 | Links strength values different from my map |
| 10 | Total number of shared links with my map |
| 20 | Number of links in my map (15 nodes) |
| 50.00% | Percent of shared links with my map |

Figure 3. Visually comparing student 4's first map with the aggregated group map (g1) with darker links revealing matching causal strength values, lighter links revealing shared links (differing in values), and light gray links revealing missing links.

The following case study illustrates how jMAP can be used to assess how differences in the quality of learner's argumentation affect changes in learner's causal diagrams. Furthermore, this paper will also demonstrate how jMAP can be used to compare causal maps between learners, identify differences between learners' maps and initial/current consensus on map links, and to initiate and structure learners' discussions in ways that might help to improve their causal maps. The case study presented in this paper presentation addressed the following research questions:

1. *What are the effects of consensus observed in initial maps on the level of consensus in subsequent maps?* When learners use jMAP to determine which causal links are shared most among everyone's initial maps, are the most commonly shared links more likely to remain in learners' subsequent maps than the less commonly shared links?

2. *What is the relationship between initial levels of consensus and level of argumentation?* Do learners engage in more argumentation when a causal link is more or less commonly shared between learners? In other words, do higher or lower levels of initial consensus trigger higher levels of argumentation?

3. *What are the effects of argumentation levels on consensus in subsequent maps?* Do high levels of argumentation lead to higher or lower levels of consensus in maps produced subsequent to group discussions/debates?

# Method

## Participants

Nineteen graduate students (8 male, 11 female) enrolled in an online course on computer-supported collaborative learning at a large southeastern university participated in this study. The participants ranged from 22 to 55 years in age, and the majority of the participants were enrolled in a Master's level program in instructional systems/design

## Procedures

The course examined factors that influence success in collaborative learning and instructional strategies associated with each factor. In week 2, learners used a Wiki webpage to share and construct a running list of factors believed to influence the level of learning or performance achieved in group assignments. Students classified and merged the proposed factors, discussed the merits of each factor, and voted on the factors believed to exert the largest influence on the outcomes of a group assignment. The votes were used to select a final list of 14 factors that learners individually organized into causal maps.

In week 3, students were presented six example maps to illustrate the desired characteristics and functions of causal maps (e.g., temporal alignment, parsimony). Students were provided the jMAP program (pre-loaded by the instructor with nodes for each of the 14 selected factors) to construct their first causal diagram (map 1). Map 1 allowed students to graphically explain their understanding of how the selected factors influence learning in collaborative settings. Using the tools in jMAP, learners connected the factors with causal links by: (a) creating each link with varying densities to reflect the perceived *strength* of the link (1 = weak, 2 = moderate, 3 = strong); and (b) selecting different types of links to reveal the level of evidentiary support (from past personal experiences) for the link. Personal maps were completed and electronically uploaded within a one-week period to receive class participation points (class participation accounted for 25% of the course grade). The maps were also used to complete a written assignment describing one's personal theory of collaborative learning (due week 4, and accounting for 10% of course grade).

Using jMAP, the instructor *aggregated* all the initial maps ($n$ = 17) submitted by students into a group matrix. The group matrix was shared with students to convey to the students the percentage of maps that possessed each causal link. The links enclosed in boxes in the right side of the figure are *common links* observed in 20% or more of the learners' maps. For example, the causal link between 'Individual Accountability' and 'Learner Motivation' was observed in 47% of learners' maps. To select this 20% cut-off criterion, the instructor ran multiple aggregations of the learner maps at different cut-off criterion until the instructor felt that a sufficient number of links were present for group discussion and to discriminate between links that were more versus less shared between learners. A corresponding group matrix was also produced in jMAP to report the mean strength values of the links observed in 20% or more of the maps. The values were highlighted by jMAP to reveal which links were present in the expert's map but missing in the group map (i.e., dark shaded cells with values = links shared and strength values match, lightly shaded with values = links shared with non-matching values, lightly shared boxes with no values = missing target links).

In week 9, learners were shown the matrix that revealed the percentage of maps (map 1) that possessed each link. Students posted messages in an online threaded discussion forum to explain the rationale and justification for each proposed causal link. Each posted explanation was labeled by learners with the tag 'EXPL' in message subject headings. Postings that questioned or challenged explanations were tagged with 'BUT.' Postings that provided additional support were tagged with 'SUPPORT.' In weeks 9 and 10, learners searched for and reported quantitative findings from empirical research into a group Wiki that could be referenced and used later to determine the instructional impact of each factor.

Students received instructions on how to use jMAP to *superimpose* their own map over the aggregated group map (Figure 2) to visually identify similarities and differences between their own maps and the collective conception of the causal relationships between factors and outcomes. For example, figure 2 reveals the similarities and differences between an individual student's first map (student #4) and the group map (g1) generated by the aggregation of all the maps produced by all students at the first time period. The course instructor used jMAP to superimpose his expert map over the group map produced at time period one (g1) and in time period two (g2) by using the control keys (ctrl-h, ctrl-j, ctrl-k) to toggle between maps g1 and g2. By using the navigational tools to toggle between the two group maps, the instructor was able to visually and quantitatively observe the progression of changes averaged across all the students' maps in order to assess the extent to which the observed changes converged towards the expert map. Jeong (2008) presents more detailed information on how to use jMAP to visualize and animate progressive changes in maps created by a select learner (or group of learners) across multiple time periods relative to a target map.

In week 10, students reviewed the discussions from week 9. Within a discussion thread for each examined link, learners posted messages to report whether they rejected or accepted the link (along with explanations). At the end of week 10, each student posted a revised causal diagram based on their analysis of the arguments presented in class discussions.

### Data Analysis

To measure the level of change in learners' maps, link frequencies from each learner's second map (*n* = 15) were aggregated to determine the percentage of maps that shared each link. Differences in the reported percentages between maps 1 and 2 were computed and presented in a matrix. Overall, the percentages in 19 of the 24 commonly shared links increased by an average of 26%. Four of these shared links changed by an average of -10.75%.

The level of critical discourse produced within each discussion on each link was determined by the number of observed EXPL-BUT, BUT-BUT, BUT-EXPL or SUPPORT, and BUT-SUPPORT exchanges. Challenges to explanations, and explanatory responses to challenges were used as a measure of critical discourse because explanations, when generated in direct response to conflicting viewpoints, have been shown to improve learning (Pressley et al., 1992). Pearson correlations between variables are presented below.

Table 4. Correlations (*n* = 24) between level of initial agreement, critical discourse, and change percent of learners sharing each causal links

| | | LevelAgree | CritDisc | %Change | Expl | But | Support | Expl-But | But-Ex/Sup | But-But | Expl-Sup |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LevelAgree | *r* | 1 | .385 | -.089 | .233 | .328 | .291 | .330 | .365 | .177 | .153 |
| | *signif* | | .063 | .679 | .272 | .118 | .168 | .115 | .079 | .409 | .476 |
| CritDiscourse | *r* | .385 | 1 | -.152 | .339 | .921 | .120 | .867 | .921 | .494 | -.135 |
| | *signif* | .063 | | .478 | .105 | .000 | .575 | .000 | .000 | .014 | .530 |
| PercentChange | *r* | -.089 | -.152 | 1 | -.058 | -.173 | .313 | -.051 | -.167 | -.219 | .386 |
| | *signif* | .679 | .478 | | .788 | .420 | .136 | .814 | .435 | .304 | .063 |

## Findings

### Effects of consensus in initial maps on level of consensus in subsequent maps.

Based on links (*n* = 24) that were observed in 20% or more of students' maps and discussed by students on the discussion board, the correlation (Table 4) between the percentage of students that shared a causal link in the first map and the average change in the percentage of students that shared the causal links was not significant ($r = -.089$, $p = .679$). The opinions of the majority did not appear to influence learners' decisions to include or exclude causal links into their revised maps. This suggests that the use of jMAP to reveal the similarities and differences between students' maps did not foster group think.

### Relationship between initial agreement and level of critical discourse.

The correlation (*n* = 24) between the percentage of students that shared a causal link in the first map and the level of critical discourse that was generated by students to exam the strength of each causal link approached statistical significance ($r = .385$, $p = .063$). The students engaged in more critical discussion over the causal links when the causal links were shared by more students rather than less students. This finding suggests that students did not simply accept or give into the status quo. Conversely, the finding also suggests that students exhibited some tendency to engage in *less* critical discussion over the causal links when the casual links were shared by *fewer* students. One possible explanation for this finding may be that the causal links shared by the fewest number of students where those that exhibited the most obvious flaws in logic and as a result, these links did not warrant much debate to omit the causal link from the causal maps.

### Effects of argumentation on changes in agreement in subsequent maps

No significant correlation was found between the level of critical discourse over each causal link and the change in the percentage of maps sharing each casual link ($r = -.152$, $p = .478$). This finding suggests that the level of critical discourse over each causal link neither increased nor decreased the percentage of students that rejected a causal link.

However, post-hoc analysis on the individual effects of each of the four types of exchanges (all of which were aggregated and used to measure the level of that critical discourse) revealed the frequency of EXPL-SUPP exchanges observed in discussions over each link were moderately and positively correlated (*r* = .386, *p* = .063) with changes in the percentage of students that shared each causal link. Supporting statements that were specifically posted in direct response to other learners' causal explanations (e.g., presenting supporting evidence, simple expression of agreement) were the types of events/exchanges that were most likely to persuade learners to adopt new links into subsequent causal maps. This finding is consistent with the findings from a previous case study in which causal link strength values were more likely to remain the same or increase in value when links were supported with evidence. Worth noting here is that the frequency of supporting statements alone in the discussions over each causal link (without regard to what messages they were posted in response to) revealed a similar correlation but of lesser statistical significance (*r* = .313, *p* = .136). This suggests that message-response exchanges as opposed to simple message frequencies alone could provide more explanatory power when analyzing the effects of critical discourse on causal understanding.

### Additional findings

To be included at the time of presentation will be the findings produced from the visual comparison of the transitional state diagrams (like Figure 1) depicting how the causal maps (or more specifically, the strength values of each causal link) changed over time resulting from the high vs. low presence of EXPL-SUPP exchanges observed in the discussions of each causal link.

## Implications

The findings in this case study illustrate how jMAP can be used to assess the impact of critical discussions or other instructional events/interventions on learners' causal understanding. When used as a research tool, jMAP provides insights into the processes of learning (e.g., causal understanding) and insights into how specific processes (e.g., EXPL-SUPP) lead to specific learning outcomes/behaviors. At the same time, this case study illustrates how jMAP can help learners work collaboratively to build and refine their causal understanding. Learners can identify similarities and differences in their causal understanding relative to others. Then they can use the differences as the starting point to discuss and explore the causal relationships.

### Directions for Future Research & Development

The findings in this case study are not conclusive given the limited sample size. Nevertheless, this study illustrates how jMAP can be used to assess how causal understanding evolves over time and how specific processes of discourse (including processes of scientific inquiry) influences causal understanding. More research is needed to identify the specific discourse processes and interventions that foster critical discourse that can trigger changes in causal links – particularly changes that converge towards the expert and/or group model.

To support future research, online discussion boards could be integrated into jMAP to automatically create discussion threads for each causal link observed in learners' causal maps, to seed discussions with learners' initial explanations, to support message tagging, and to compile and report scores that measure certain qualities observed in the group discussions for any given set of causal links. Such a system could be used by instructors to assess not only the quality of learners' causal maps and understanding, but also the quality of learners' discourse and its impact on their causal understanding. Additional functions can be added to jMAP to recognize nodes that are indirectly linked via mediating nodes to fully account for observed differences between learner and expert maps. Another useful function would be one that can identify and measure to what extent and in what temporal direction changes in causal links propagate subsequent changes in adjacent links – a measure that could be used to determine to what extent learners apply a systematic approach to break down the causal relationships. To examine this issue in more detail, a function can be added to jMAP that captures and logs every action performed in jMAP as learners construct their maps.

In addition, refinements to the jMAP user interface will be necessary to make map construction easier, more intuitive, and less time consuming if systems like jMAP are to be used in school-based applications – particularly for learners at younger ages. Instructions and guidance on how to conceptualize a coherent causal map/model (e.g., temporal flow, parsimony) should be embedded directly into the jMAP interface to assist learners that lack the skills needed to construct a causal map. Other useful functions to add to jMAP include: swap and change the target map so that one can compare different combinations of maps (individual, collective group, expert); show the percentage of students' maps that possess each target link by varying the density of each link to reflect the observed percentages; and select links to include in the aggregate/collective group map based on link frequencies that are significantly higher than the expected frequencies based on a user-selected critical z-score.

## References


Doyle, J., M. Radzicki, et al. (2007). Measuring change in mental models of complex systems. In *Complex Decision Making: Theory and Practice*. H. Qudrat-Ullah, J. M. Spector and P. I. Davidsen, Springer-Verlag: 269-293.

Ifenthaler, D., & Seel, N. M. (2005). The measurement of change. Learning-dependent progression of mental models. *Technology, Instruction, Cognition and Learning, 2*(4), 317-336.

Ifenthaler, D., M. Iskandaria, et al. (2008). Tracking the development of cognitive structures over time. Paper presented at the American Educational Research Association 2008 conference, New York, NY.

Jeong, A. (2005a). Discussion analysis tool. Retrieved May 2009 http://myweb.fsu.edu/ajeong/dat.

Jeong, A. (2005b). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education, 26*(3), 367-383.

Jeong, A. (2008). jMAP. Retrieved July 24, 2008 http://jmap.wikispaces.com

Nakayama, V. K. & Liao, J. (2005). An outline of approaches to analyzing the behavior of causal maps. In V. K. Nakayama and D. J. Armstrong (Eds.), *Causal mapping for research in information technology.* Hershey PA, Idea Group Publishing: 368-377.

Nesbit, J., & Adesope, O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*(3): 413-448.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist*, *27*(1), 91-109.

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition and Learning, 1*(1), 59-85.

Shute, V. J., & Jeong, A. C., & Zapata-Rivera, D. (in press). Using flexible belief networks to assess mental models. In B. B. Lockee, L. Yamagata-Lynch, and J. M. Spector (Eds.), *Instructional Design for Complex Learning*. New York, NY: Springer.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building).* Syracuse, NY: Syracuse University.

# THE EPISTEMOGRAPHY OF URBAN AND REGIONAL PLANNING 912: APPROPRIATION IN THE FACE OF RESISTANCE

Elizabeth Bagley and David Williamson Shaffer
University of Wisconsin-Madison, 1025 West Johnson Street 1078 Educational Sciences, Madison, WI, 53706
Email: easowatzke@wisc.edu, dws@education.wisc.edu

**Abstract:** Preparing citizens to address the complex problems inherent in cities requires our changing society to embrace a new kind of education. One way to train people to think about complex problems is to identify and study how professionals who think in those ways develop their *epistemic frame*. In this paper, we examine one of the ways urban planners master and appropriate relevant expertise through an ethnographic study of an urban planning practicum. Specifically, we use a new method called *epistemic network analysis* to look at presentation feedback sessions during two weeks of the practicum to explore emergent relationships between the teacher's planning expertise and the students' expertise. The results of this study indicate that epistemic network analysis offers a technique for analyzing the kinds of situated understanding that result from sociocultural learning and for observing the translation of pedagogy into practice in various types of learning environments.

## Major issue(s) addressed

Measuring learning in a practicum environment can be challenging, and a growing body of research suggests that a new method called epistemic network analysis (Shaffer et al., 2009) can inform our understanding of how professionals-in-training learn in a practicum environment. According to Donald Schön (1983; 1987), a practicum environment is explicitly designed to forge the links between knowing and doing that are central to the reflective practice of a profession. In a practicum, novices are initiated into a professional community of practice and extend their knowledge through tackling complex problems.

John Friedmann (1973) argues that our changing society requires a new kind of education where knowledge is extended and people are trained to think about and address the complex problems inherent in cities. One group of professionals tasked with addressing several of these challenges is urban planners, and studying how professional urban planners learn in a practicum environment can help us better understand how to train people to address complex problems.

In this paper, we examine one of the ways urban planners develop expertise through an ethnographic study of Urban and Regional Planning (URPL) 912, a graduate level practicum at a large Midwestern university. The main goal of the study was to explore the learning processes experienced by the 20 graduate students in the practicum. The students were guided in the production of a site plan for a developing area by a planner with 34 years of planning experience. In the study, we used epistemic network analysis to examine the presentation feedback sessions during weeks four and five and to explore emergent relationships between the teacher's planning expertise and the students' expertise.

## Potential significance of the work

The results of this study have the potential to influence the future design of professional practicum environments as well as the broader landscape of education.

## Theory

A major goal for educators is creating instructional contexts in which skills are both mastered and appropriated (Herrenkohl & Wertsch, 1999). Mastery and appropriation, according to Wertsch and Polman (2001), are part of mediated action—human action that is fundamentally characterized by a tension between active agents and the cultural tools they use to carry out action. Wertsch and Polman (2001) define mediated action as forms of action such as speaking, reasoning, and calculating that inherently involve agents actively using cultural tools. Herrenkohl and Wertsch (1999) stress that mastery of a cultural tool involves having the skill to use a cultural tool effectively, "knowing how" as opposed to "knowing that," or in other words, procedural versus declarative knowledge. In contrast, appropriation focuses on an agent's tendency to use a cultural tool, which can be distinct from the level of mastery involved. Using Bakhtin (1981), Herrenkohl and Wertsch (1999) claim that appropriation means to adopt, imitate, or pick up someone else's accent. Thus, appropriation is a process of making something, such as a historical narrative, one's own.

High levels of mastery are frequently associated with appropriation; however, some forms of mediated action are characterized by mastery but not appropriation of a cultural tool. Bakhtin argues that cultural tools are often not easily and smoothly appropriated, and that an agent may use a cultural tool but with a feeling of resistance or even outright rejection. When such resistance grows sufficiently strong, the agent may refuse to use the cultural tool altogether (Wertsch, 1998). However, Wertsch (1998) writes, "it has become increasingly clear that interactional contexts involving resistance and rhetorical opposition may provide some of the most productive settings for developing mastery and appropriation of cultural tools" (p. 182).

Herrenkohl and Wertsch (1999) believe that one of the most effective ways to foster the appropriation, and not just the mastery of cultural tools is to coordinate these cultural tools with sociocognitive roles. They claim that sociocognitive roles can be understood in terms of rights and responsibilities, where people have opportunities to exercise their rights as a way of being responsible to their community. Herrenkohl and Wertsch (1999) offer the example wherein a building inspector exercises her right to stop construction on a building because the contractor is suspected of using sub-par materials. In this example, the inspector is exercising a right in the context of her responsibility to protect public safety. Or, put simply, she is performing her job.

Herrenkohl and Wertsch (1999) propose that by promoting the idea of "doing one's job" and emphasizing the responsibilities to one's community and the set of rights that accompany those responsibilities, students will practice skills important to the sociocognitive role and begin to master and appropriate them. Unfortunately for educators hoping to introduce sociocognitive roles into their classrooms, Herrenkohl and Wertsch do not outline a specific process for creating the sociocognitive roles or offer suggestions about the types of skills that could be mastered or appropriated through students assuming roles.

Schön (1983; 1987) argues that, in most professions, people begin to master and appropriate skills in professional practicum experiences. In a professional practicum, novices engage in simulations of professional work. Their work is guided by repeatedly taking action and explicitly reflecting on that action with peers and mentors, what Schön refers to as reflection-on-action. The process of explicit reflection-on-action allows one to look back on a completed task or process to consider the implications and consequences of actions. Schön (1983; 1987) argues that the goal of the professional practicum is to bind action and reflection together to produce professional expertise particular to each profession.

Extending Schön, Shaffer (2004a; 2004b; 2006) argues that a professional practicum is a key step to developing the epistemic frame—or the ways of knowing, of deciding what is worth knowing, and of adding to the collective body of knowledge and understanding—of a particular community of practice. In a practicum environment, experienced mentors explicitly reflect-on-action as a way to model the epistemic frame of a profession. For novices, iterative cycles of action and explicit reflection-on-action with peers and mentors bind together the elements of the epistemic frame—the skills, knowledge, values, identity, and epistemology—that an individual takes on as a member of a community of practice. This collection of cultural tools forms the epistemic frame of the community, which, once appropriated, can be used when an individual approaches a situation from the point of view (or in the role) of a member of the community (Shaffer 2004a; 2004b; 2005; 2006).

Thus, Schön (1983; 1987) and Shaffer (2006) have a model of learning in a professional practicum that involves iterative cycles of action, explicit reflection-on-action, and the construction of a profession-specific epistemic frame. Their model extends Herrenkohl and Wertsch's (1999) assertion that coordinating cultural tools with sociocognitive roles can lead to both mastery and appropriation. Specifically, Schön and Shaffer move beyond a conversation about rights and responsibilities and into a discussion about designing professional practica and building epistemic frames. However, Schön and Shaffer's model does not address the role resistance plays in the mastery and appropriation of an epistemic frame.

While it is possible, and often quite important, to analyze how well students and others have mastered a cultural tool, such analyses can be quite limited in that they do not consider all of the complexities in the relationship between agents and the cultural tools they use (Wertsch, 1998). Thus, measuring appropriation in a practicum setting requires analysis of the process over time to see if there are instances of resistance that are inhibiting the appropriation of the epistemic frame. One way to analyze those components is through an epistemography, an analysis of the structure of a professional practicum through the lens of epistemic frames where one can examine the kinds of action and reflection-on-action that develop the epistemic frame of a profession (Shaffer, 2005; Svarovsky, 2006; Hatfield, 2008). An epistemography allows one to see learning principles at work and recognize some features of the practicum as being more essential than others in developing the professional epistemic frame. However, as Wertsch (1998) noted, the relationships between agents and their use of cultural tools are complex, and traditional statistical methods do not account for the complexities.

The kinds of professional understanding that a practicum develops are complex because they are not merely a collection of disconnected skills and knowledge. Rather, the power of an epistemic frame is in the connections among its parts, a network which consists of relationships among conceptual, practical, moral, personal, and epistemological parts (Shaffer et al., 2009). Thus, analytical methods such as social network analysis provide a robust set of tools for representing networks of relationships, including complex and dynamic relationships of the kind that characterize epistemic frames (Shaffer et al., 2009). In social networks, individuals are considered nodes in the network and relationships between individuals are represented as arcs or links between nodes (Haythornthwaite, 1996). For example, a social network analysis of an urban planning practicum might examine the relationships among students and the teacher throughout class meetings. Within each class session, different configurations might emerge as old friends connected, new friendships emerged, and different team projects occurred. The amount of time individuals spend with each other could be taken as a proxy for the strength of their relationship by analyzing the different connections among and between nodes and links. That type of analysis would provide a quantifiable way of comparing social relationships across time and a means for better understanding the informal information flows that supplement the formal practicum curriculum.

However, as Shaffer (2009) argues, social network analysis was developed to provide insight into relationships among and between individuals and groups, rather than relationships within the conceptual, practical, moral, and epistemological world of an individual. Therefore, building on social network analysis, Shaffer (2009) has developed epistemic network analysis, a computational modeling technique for the development of epistemic frames.

Epistemic network analysis is based on two key concepts: (1) that thinking can be characterized by the application of an epistemic frame composed of the linkages between professional skills, knowledge, identity, values, and epistemology; and (2) that the development of professional thinking can be quantified, analyzed, and visualized with a dynamic network model of the developing epistemic frame (Shaffer et al., 2009). Epistemic network analysis has been used to trace frame development in elementary and middle school students during epistemic games based on engineering and urban planning (Nulty & Shaffer, 2008; Nash & Shaffer, 2008). A preliminary re-analysis of qualitative data collected on a science journalism practicum suggests that epistemic network analysis will be a useful tool for analyzing epistemic frame development in professional practica (Hatfield, 2008).

This study extends the ideas of Wertsch, Schön, and Shaffer by examining the relationships between appropriation, resistance, reflection-on-action, and epistemic frames in a professional planning practicum. The aim of this study—the epistemography of URPL 912—is to uncover the learning process within a graduate urban planning practicum. In particular, we investigate how one teacher communicated his urban planning epistemic frame in the face of resistance by describing the students' initial resistance to the teacher's frame, the teacher's explicit reflection-on-action, and the students' appropriation of the teacher's frame. We then use epistemic network analysis to examine the teacher's role in the students' epistemic frame development by tracking how specific features and events in the practicum led to significant changes in frame development. We argue that epistemic network analysis can provide a computational model of the extent to which participants appropriated the ways of knowing, being, talking, and acting that characterize a particular community of practice. Finally, we discuss how the results of this study may contribute to the design of reflective learning environments and experiences that promote the development of the next generation of urban citizens.

## Methods

Urban and Regional Planning (URPL) 912 was a three credit course that met once a week for 14 weeks for approximately three hours. We chose to study URPL 912 because the course gives graduate students an opportunity to work in a practicum setting and is a prerequisite to entering the professional field of planning. The teacher was a professional planner with 34 years of planning experience across the United States.

In the course, 20 graduate students from the URPL masters program prepared a site plan for a developing area of approximately 3,000 acres on the northeast edge of Madison, Wisconsin. In the syllabus, the teacher wrote that he expected the students to "read the landscape" and expand upon the City's draft neighborhood plan for the area. Most of the class sessions began with the teacher's lecture, class discussion, or a professional planner guest speaker for the first hour and team work time for the remaining two hours.

Classroom data were collected in digital audio recordings and supplemented with field notes. Recordings were transcribed to provide a detailed record of interactions, and field notes were used to capture meaningful non-verbal aspects of the context and to supplement the transcripts. No information on specific career plans was collected; however, several students mentioned plans to pursue planning in city departments, non-profit groups, and state agencies. No other demographic information was collected about the students.

The data were segmented into interactive units which were defined as strips of activity with a consistent interactional structure and topical focus. For example, if the class started discussing the capacity of a proposed wastewater treatment plant and then switched to discussing the location of bike and pedestrian paths, the switch in discourse topic would indicate two separate interactive units. If an interactive unit represented more than one category, it was coded for all applicable codes. Within each interactive unit, the students' comments were coded cumulatively instead of individually in order to compare the students' cumulative frame to the teacher's frame.

This study's goal was to observe students learning to become planners through participation in a practicum. To capture interactions between the expert teacher and the novice students, we decided to focus on the communication between the teacher and the students during presentation feedback sessions. Presentation feedback sessions were occasions for the teacher and the students to offer feedback on information teams collected and for the teacher to explicitly reflect-on-action. Those sessions occurred in four classes throughout the semester: weeks 4, 5, 11, and 13. We chose to analyze data from the presentation feedback in weeks four and five because the feedback given during those weeks focused on the information needed to create successful site plans whereas the feedback given in weeks 11 and 13 was more focused on the logistics of preparing for the final presentations. The specific activities in weeks four and five are outlined in more detail in the results section.

We used the teacher's epistemic frame in week four as the comparative model for the students' cumulative epistemic frame in both week four and week five because we were interested in seeing if the epistemic frame the teacher used during week four influenced the students' epistemic frame in week five. Additionally, after giving his initial lecture in week five, the teacher did not contribute as much as in week four. To measure the teacher's contribution in weeks four and five, interactive units were coded for the presence of the teacher's comments.

Table 1: Analytic codes applied to segmented interactive units for qualitative data analysis.

| Code | Description | Example |
|---|---|---|
| Resistance | Student references to their conception of how the planning process should progress in a way that was contrary to the teacher's conception of how the process should move forward | …I sort of got the sense that they [the City] want us to deliver to them a set of policy recommendations and other higher level stuff to help them move this process along rather than our own design. |
| Reflection-on-action | References where the teacher looked back on a completed task or process to consider the implications and consequences of actions | …I'm just saying that when you look at the land use pattern, based on uses like that quarry, there's real limitations on residential in a large part of the area...Maybe this needs to be a place with a real employment center instead of just a bedroom community. I'm not saying that you don't have residential, and I'm not even saying you necessarily start in one place or the other, but I'm saying that it's gotta be in the thought process here. |
| Skill of reading a landscape (S/L) | Ability to use the landscape to inform the planning process | …I that it's very important that we step back and say, "How does this landscape speak to us?" Rather than letting someone else give it to us. |
| Skill of suggesting alternatives (S/A) | Ability to use a specific strategy or an alternative way to approach creating a plan | Shoehorning residential onto the North end won't work. Let's see if we can put some jobs up there so that the people who live further south, where it's easier to do residential development have a place to go that's a mile away or a mile and a half away instead of coming down to Madison. With that, make it a more sustainable community. |
| Skill of questioning assumptions (S/QA) | Ability to identify the assumptions the students made in their analyses | You guys all have to make some recommendations, but I don't want us to go into this without making sure that we are comfortable with the assumptions they [the City] are operating under…I'm not comfortable with all of the assumptions they are operating under and I think that their assumptions are no longer evidence-based… |
| Knowledge of additional information (K/AI) | References to specific information that might be useful for creating the site plan. | This is what we think is important to identify: Property values and who owns it to see if there's any correlation there. Target areas for potential development areas etc... Changes in zoning with different incentives for developers, transfer of development rights. Existing viewshed protection… |
| Knowledge of past process (K/P) | References to the process used by the City of Madison to create a plan for the same redevelopment area. | We saw a map of what they [the City] have in mind, and they already have land uses plotted out. They are presenting that to the mayor in the next few weeks. In that land use map, they have mixed use housing and TODs [transit oriented developments] and lower density housing, and the majority of it is also going to be lower density acreage. |

| | | |
|---|---|---|
| Value of serving the public interest (V/PI) | Stating that the needs of people affected by the planning process are important to urban planning practice | The developer has to be involved in this association. It's not optional. They have to be part of the deal, and you are going to have to figure out what the City of Madison would say to one or more private land owners. |
| Epistemic statement about stakeholders' desires (E/SD) | Justifying decisions based on how participants thought a particular stakeholder group would respond. | …Businesses that might be developing through the university or incubator or something. They are going to need production and assembly facilities. They are going to need distribution facilities. And that might not be stuff they can get either on campus or in that incubator. This might be an ideal spot for them… |
| Epistemic statement about principles of good urban form (E/UF) | Justifying decisions based on the principles of good urban form | The city is planning the East Wash. build out in terms of employment…which means that it would have to put itself on the periphery because of the land loss. They could have an office near their production facilities. We see that as a potential benefit. |

The result of the coded segments was a database of interactive units showing the presence of the teacher's comments, resistance, reflection-on-action, and epistemic frame elements. The relationships among these different components were then analyzed using epistemic network analysis to identify salient themes.

Epistemic network analysis adapts the framework of social network analysis for use with cognitive, rather than social, elements (Hatfield, 2008; Shaffer et al., 2009). As discussed in the background section, Shaffer (2009) has developed epistemic network analysis, a computational modeling technique for the development of epistemic frames. For details on the computations involved in epistemic network analysis, please see Shaffer (2009).

Once an epistemic frame is represented as a series of cumulative adjacency matrices showing the strength of association between each pair of frame elements for a given participant in the data set, the characteristics of the network can be quantified using concepts from social network analysis, such as network density and centrality of individual nodes (Shaffer et al., 2009). The overall structure of an epistemic frame can then be quantified by computing the relative centrality of each node: the square root of the sum of squares of its associations with its neighbors expressed as a percentage of the weight of the heaviest node in the network.

For this study, the relative centrality of each frame element and sub-element at the final time slice of the presentation feedback activity was calculated in order to compare the cumulative students' frame in weeks four and five to the teacher's modeled frame in week four. Calculating the relative centrality exposed which frame elements were further or closer to the center of the epistemic network (relative centrality values closer to 100) since epistemic frames consist of elements linked together with some elements more central than others. Using only the final time slice of relative centrality offered the most accurate picture of the students' cumulative frame development during weeks four and five since relative centrality is a cumulative measure of the changes in centrality to the epistemic network graph.

The frame similarity index (FSI) is an extension of epistemic network analysis. In this study, the FSI was calculated using the relative centrality which allowed comparisons between the students' frame development in weeks four and five and the frame modeled by the teacher in week four. The FSI was computed by using the difference between the cumulative students' final relative centrality calculations for each frame element in weeks four and five and the teacher's final relative centrality calculations for each frame element in week four. The absolute value of the difference of the values was then calculated. In order to make claims about skills, knowledge, values, and epistemology as a whole, the average of the frame sub-elements (e.g. E/SD and E/UF) was computed, and the results were graphed as cumulative frame elements (e.g. epistemology instead of E/SD and E/UF).

The FSI affords us the ability to use qualitative data to compare the development of complex thinking quantitatively. The FSI can be visualized using a radar plot where the optimal frame is represented as the origin and the developing frame is shown as something. This type of representation allows us to visualize the movement of novices' epistemic frame development in relation to the expert's epistemic frame. In the radar plot below, the teacher's frame is represented as the origin, and the closer the students' cumulative FSI is to zero, the more closely the students' frame resembles the frame the teacher modeled in week four.

## Results

We describe our observations of URPL 912 in three parts below. First, we identify and describe the interactive units in which the students resisted the teacher during weeks four and five. Next, we identify and describe the interactive units in which the teacher reflected-on-action during weeks four and five. Finally, we describe the changes in the students' cumulative epistemic frame from week four to week five.

At the beginning of the semester, students in URPL 912 worked in teams to gather information about the redevelopment site. During the first class session, they learned that the City of Madison had been working on a plan for the site for three years, and the students were expected to "read the landscape" and expand upon the City's draft neighborhood plan for the area. The students continued to learn about the site through the teacher, guest speakers from the City of Madison planning department, out of class site visits and meetings with city officials, and targeted internet research. The teams were expected to present their initial findings during week four and their more specific findings during week five.

While presenting and giving feedback during week four, the students referred to the approach the City of Madison was using for the redevelopment site. When the teacher suggested alternative approaches that were contrary to the City's approach, the students began to resist his suggestions. Overall, in week four, 3 of the 11 segmented interactive units in the presentation feedback activity were coded for the students resisting the teacher's ideas, accounting for 54% of the time when both the students and the teacher were talking about the same topic. In contrast, in week five, the students did not resist the teacher's suggestions in any of the seven segmented interactive units in the presentation feedback activity when both the students and the teacher were talking about the same topic.

During the presentation feedback activities, the teacher often explicitly reflected on the students' findings, gave suggestions for additional information they could gather, and shared anecdotes about how the problems they were facing were similar to problems he had faced with previous projects. In week four, 11 of the 12 total segmented interactive units were coded for the presence of the teacher's comments, and 8 of the 11 segments during which the teacher spoke were coded for him reflecting-on-action. By reflecting-on-action, the teacher specifically pushed the students to question the City's assumptions and to consider how the population projections would affect traffic, jobs, and the overall development trajectory. By explicitly questioning the City's assumptions and offering suggestions about how to deal with multiple possibilities, the teacher spoke directly to the students' resistance in week four and strengthened the case for having the students' use his approach rather than the City's approach.

Unlike week four, in week five, the teacher started the class session by giving a lecture. During his lecture, the teacher reflected-on-action and explicitly addressed the students' resistance from week four. He reflected on the actions the students took in week four and referred to his experience as a planner in order to address the students' anxiety about using his approach instead of the more familiar City's approach. Specifically, the teacher suggested that taking time to gather information early in the process would have positive implications for the final site plan. To address their anxieties, he encouragingly reflected on the work the students completed in week four, told them he was expecting to be pleased in week five, and assured them that though they were at a slow part in the process, they were on course. To attend to the students' concern that his approach would not relate to the work the City had already accomplished, the teacher suggested that the students: "…Let this piece of land speak to us [because]…If we try to decide what it's going to look like before then, what you're going to end up with is exactly what you don't want to end up with which is something that doesn't relate."

Following the teacher's lecture, the students presented the information they gathered about the site and gave feedback on the information presented. During the presentation feedback, 8 of the 14 total segments were coded for the presence of the teacher's comments, and in one of the eight segments, the teacher was the only person speaking. Two of the seven segments wherein the teacher spoke were coded for the teacher reflecting on action. During the presentation feedback in week five, a student asked about the City's assumptions, and the teacher reflected-on-action by explicitly considering the implications and consequences that information would have for the recommendations they could make in their final site plans. By explicitly reflecting-on-action and addressing the student's resistance before they rejected his process outright, these data suggest that the teacher created a space where the students could begin to appropriate the epistemic frame he modeled in week four.

Using epistemic network analysis, Figure 1 provides a summary representation of the frame similarity index (FSI) of the difference between the students' cumulative epistemic frames in weeks four and five and the teacher's modeled epistemic frame in week four, which is represented by the origin of the radar plot. The total FSI was 165.8 in week four, but in week five, the FSI decreased to 90.3 suggesting that the students' frame became more similar to the frame the teacher modeled in week four. There was not a significant change in value development from week four to week five for the students; however, skills, knowledge, and epistemology frame elements began to look more like the teacher's modeled frame.

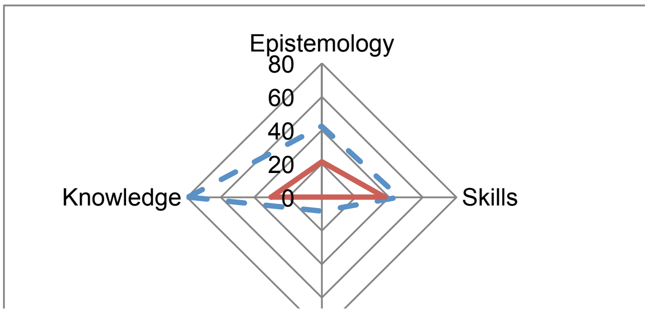



Figure 1. The origin of the radar plot represents the teacher's frame, and as the cumulative students' FSI moves closer to zero, their frame begins to look more like the frame the teacher modeled in week four.

By separating the epistemic frame elements into sub-elements, a more complete picture emerged about which specific sub-elements became more or less central to the students' epistemic frame from week four to week five (Table 2). As discussed above, for the teacher, V/PI, S/QA, S/Land, K/AI, and E/UF were the most central sub-elements in his epistemic frame in week four while the most central sub-elements in the students' epistemic frame were V/PI, E/SD, and K/P. In week five, instead of having a strong central core consisting of V/PI, E/SD, and K/P, the students exhibited a new configuration of their epistemic frame which looked more like the teacher's and included sub-elements such as S/Land, S/Alt, K/AI, and E/UF increasing in centrality. The order of centrality of frame elements also changed from week four to week five.

According to Shaffer (2009), the relative centrality of a node within a network represents the extent to which the node is or is not part of the dense central core of the network. Thus, though some of the sub-elements became more central to the students' epistemic frame, the relative centrality values of the sub-elements in the students' frame in week five were consistently lower than the teacher's values. Presumably, the teacher had higher relative centrality values for frame elements in week four because his 34 years of planning experience necessitated that his epistemic frame start out more richly interconnected than the students' epistemic frame. Therefore, the 'looseness' of the students' epistemic frame may be due to the students beginning to appropriate the sub-elements, and since the dense core is central to the strength of the epistemic frame, their epistemic frame will likely strengthen over time.

Table 2: Cumulative relative centrality calculations for eight epistemic frame sub-elements for the students in weeks four and five and the teacher in week four.

| | V/PI | S/Alt | S/Land | S/QA | K/AI | K/P | E/SD | E/UF |
|---|---|---|---|---|---|---|---|---|
| Students Week Four | 91.65 | 44.72 | 0 | 60 | 28.28 | 91.65 | 100 | 44.72 |
| Students Week Five | 100 | 73.38 | 55.47 | 55.47 | 83.21 | 48.04 | 55.47 | 55.47 |
| Teacher Week Four | 100 | 46.71 | 96.30 | 100 | 95.35 | 0 | 61.79 | 91.45 |

These results suggest that examining changes in epistemic frame development across time was a useful way of seeing the process of appropriation in the midst of student resistance in URPL 912. Through explicit reflection-on-action, it appears that the teacher addressed the students' resistance, enabling them to appropriate his epistemic frame.

## Conclusions and implications

This study extends Wertsch's work on resistance and appropriation. Wertsch (1998) writes about the productive role of resistance in the process of appropriation, and in week four, there was student resistance. However, looking across weeks four and five shows that the students' resistance subsided, and the bridge between the students resisting and not resisting was the teacher's lecture. The teacher's lecture was essentially an explicit reflection on the different frames held by the teacher and the students and provided a map of the professional vision of the planning practice. However, it seems unlikely that his lecture immediately helped the students understand the epistemic frame of planners in a new light and enabled them to put their new knowledge into practice in their presentation feedback. Therefore, the students must have started appropriating aspects of the teacher's epistemic frame in week four despite their resistance.

Thus, these results also suggest that the kind of reflective mentoring that Schön and Shaffer describe in professional practicum settings accomplishes the task of helping students appropriate a new frame in the face of resistance. Specifically, the results of this study indicate that identifying practicum activities that evoke evidence about certain aspects of an epistemic frame will provide valuable information for designing effective practicum environments and learning in general. For example, practitioners thinking about ways to enhance their practice might consider including iterative cycles of action and reflection-on-action which may lead to appropriation and mastery.

This study demonstrates that epistemic network analysis can be a productive way of tracking how specific interactions within learning environments lead to significant changes in cognitive development. Building on initial work (Hatfield, 2008; Shaffer et al., 2009), this study adds frame similarity index to epistemic network analysis' set of techniques, in this case to compare the students' frame development in weeks four and five to the frame modeled by the teacher in week four. Specifically, the differences between the students' and the teacher's relative centrality values suggest that indices from epistemic network analysis can be useful for group comparisons and experimental studies of interventions. In other words, epistemic network analysis provides a computational model of the process and extent to which participants appropriated the ways of knowing, being, talking, and acting that characterize a particular community of practice. Thus, epistemic network analysis offers a technique for analyzing the kinds of situated understanding that result from sociocultural learning.

This study's findings can expand epistemic network analysis to provide a computational model of the extent to which participants appropriate a professional epistemic frame in the face of resistance with the help of a mentor's explicit reflection-on-action. Thus, epistemic network analysis points towards a promising new way of observing the translation of pedagogy into practice in various types of learning environments. These findings—and future studies investigating reflective practica and the development of epistemic frames—can shed light on how to better prepare citizens to think about and address the complex problems inherent in cities because as John Friedmann (1973) wrote, "the reconstruction of society must begin with man's re-education."

## References


Bakhtin, M. M. (1981). *The dialogic imagination: Four essays by M. M. Bakhtin*. Austin, TX: University of Texas Press.

Friedmann, J. (1973). *Retracking America: A Theory of Transactive Planning*. Garden City, NY: Anchor Press.

Hatfield, D. L. (2008). Informing Expertise: Designing epistemic games based on civic professions: University of Wisconsin-Madison

Haythornthwaite, C. (1996). Social network analysis: an approach and technique for the study of information exchange. *Library and Information Science Review, 18*, 323-342.

Herrenkohl, L., & Wertsch, J. V. (1999). The Use of Cultural Tools: Mastery and Appropriation. In I. E. Sigel (Ed.), *Development of Mental Representation: Theories and Applications* (pp. 415-435). Mahwah, NJ: Lawrench Erlbaum Associates.

Nash, P., & Shaffer, D. W. (2008). *Player-mentor interactions in an epistemic game: A preliminary analysis*, *International Conference of the Learning Sciences (ICLS)*. Utrecht, Netherlands.

Nulty, A., & Shaffer, D. W. (2008). *Digital Zoo: The effects of mentoring on young engineers*, *International Conference of the Learning Sciences (ICLS)*. Utrecht, Netherlands.

Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Schon, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.

Shaffer, D. W. (2004a). Epistemic frames and islands of expertise: Learning from infusion experiences. In Y. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon & F. Herrera (Eds.), *Proceedings of the Sixth International Conference of the Learning Sciences* (pp. 473-480). Mahwah, NJ: Erlbaum.

Shaffer, D. W. (2004b). Pedagogical praxis: The professions as models for post-industrial education. *Teachers College Record, 106*(7), 1401-1421.

Shaffer, D. W. (2005b). Epistemography and the participant structures of a professional practicum: A story behind the story of Journalism 828: University of Wisconsin-Madison, Wisconsin Center for Education Research.

Shaffer, D. W. (2006). *How Computer Games Help Children Learn*. New York: Palgrave.

Shaffer, DW, Hatfield, D, Svarovsky, GN, Nash, P, Nulty, A, Bagley, E, Franke, K, Rupp, AA, Mislevy, R (2009). Epistemic Network Analysis: A prototype for 21st Century assessment of learning. *The International Journal of Learning and Media*. 1(2), 33-53.

Svarovsky, G. N., & Shaffer, D. W. (2006). *The Epistemography of BME 201: Building engineers, one participant structure at a time*: University of Wisconsin-Madison.

Wertsch, J. V. (1998). *Mind as action*. New York: Oxford University Press.

Wertsch, J. V., & Polman, J. I. (2001). The Intuitive Mind and Knowledge about History. In B. Torff & R. J. Sternberg (Eds.), *Understanding and Teaching the Intuitive Mind: Student and Teacher Learning* (pp. 57-72): Lawrence Erlbaum Associates.

# Assessing the Development of Expertise in an Historical-Based Science: The Case of Integrative Archeology

Inbal Flash-Gvili, Jeff Dodick, Science Teaching Center, The Hebrew University of Jerusalem, Israel
finbal@cc.huji.ac.il, jdodick@vms.huji.ac.il

Developing inquiry-based skills in science students is a key focus of science education. This study examines the process whereby such skills are acquired by exploring the case of graduate training in the discipline of integrative archeology. To do this, we focus on the questions that the students posed during their field research. Our results show that some students attempt to bypass an initial phase of their research by trying to solve the final goal of reconstructing human behaviors, rather than focusing on material remnants that make such reconstruction possible. In the field, this was observed when the students attempted to reconstruct events via their own logic, rather than using the technical tools at their disposal to analyze the site's (chemical) properties. However, as their expertise grows, the students learn about the importance of foundational questions; this is reflected in a change in the frequencies of their questions, which in turn reflects a change in their research strategies.

## Introduction

Scientific inquiry can be defined as "the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work" (NRC, 1996 p. 23, 2000 p.1); thus, the NRC defines inquiry-based learning as a process where "students actively develop their understanding of science by combining scientific knowledge with reasoning and thinking skills" (NRC, 1996 p. 2). The same source also claims that: "Inquiry into authentic questions generated from student experiences is the central strategy for teaching science" (NRC, 1996 p. 31). These definitions create a picture of a type of learning strategy, which should reflect the nature of science via scientific inquiry. Yet, the question remains as to what exactly are 'the correct' characteristics of scientific inquiry in practice?

A simple model of scientific inquiry is based on *the scientific method*. Broadly speaking, it is described as a process in which the following events occur in sequential order: (1) observing, (2) developing a question, (3) developing a hypothesis, (4) conducting an experiment, (5) analyzing data, and (6) stating conclusions. This process continues as new questions are generated based on previous findings. This model, with different variations is repeated in many of the US standards documents (NRC, 1996, 2000).

However, since its development, this model, with its rigid series of stages, has been criticized by philosophers and science educators, who have produced alternative lists of process skills that were considered more flexible. An example of such an alternative list is presented in the *Science: A Process Approach* program (AAAS, 1967). One example of an alternative process skill is controlling variables, a strategy which has received enormous attention in research on the development of scientific reasoning (Kuhn and Dean, 2005). Still, these lists are problematic as they define *the scientific method* as a coherent entity with regards to different scientific disciplines (such as geology or physics). This problem is reflected in the way scientific inquiry is defined by the NRC (1996; 2000). From one side, it talks about multiple methods; yet, concurrently it adheres to a single scientific method. Thus, if one of our science education goals is to create future experts by enhancing their ability to acquire scientific inquiry skills, we need to teach in a much more authentic way that reflects the methodological diversity of the sciences. As Sternberg (2003) suggests:

> If we wish to teach and identify expert students, we need to identify expertise in a way that is closely aligned with the way experts are identified in the disciplines students study. For starters, this means having students do tasks, or at least meaningful simulations, that experts do in the various disciplines. Second, it means teaching them to think in ways experts do when they perform these tasks.

Following Sternberg's (2003) recommendations, if we want to understand the development of expertise, we must first characterize the tasks that experts do in different disciplines, as well as the inquiry skills that they apply while engaged in those tasks. One research method that has been used to understand expert inquiry skills in the sciences is by studying the practices of scientists while they pursue their research agenda in real time (the *in vivo* method). Thus, in this study, we ask the following question: how do inquiry skills develop within a team of graduate student researchers in the *historical-based* science of integrative archeology? To answer this question we followed these students as they conducted their research.

There are many instruments for assessing the development of scientific inquiry skills, and in this study we used (among other sources of data) the scientific questions our research subjects generated about their field-based work. Students' ability to ask high-level questions about scientific phenomena has been shown to be a good indicator of science learning in students from the high school to university level (Brill & Yarden, 2003; Chin, 2004; Hofstein et al, 2005; Marbach-Ad & Sokolov, 2000). However, such question-asking must be put in context; to do this, we will also describe the research process of the team as a whole, as well as add important observations from the field and our interviews with the team members which further validate this analysis.

### Understanding the Development of Scientific Expertise

Our understanding of what it means to be an expert in a domain has grown ever since DeGroot's (1946) classical studies of chess players. Thus, Chi and Glaser (1988) described experts as excellent performers who have superior short and long-term memory and represent problems in deeper, more principled way than novices who tend to build superficial representations. Concordantly, it was found that experts spend more time on constructing a good problem representation, while novices applied a suboptimal trial and error strategy (Van Gog et al., 2005).

Many theories were developed to explain these differences in performance (e.g.: Ericsson et al, 1993; Gobet and Simon, 1996). One such theory, the Model of Domain Learning (MDL) suggested by Alexander (1997), is important to our work because it portrays the nature of developing expertise in authentic academic domains rather than extracting it from carefully chosen problems in diverse nonacademic domains. In contrast to the traditional expert-novice theories, this model assumes that there are no sharp contrasts between experts and novices; rather, it is a graded transition toward expertise. It also considers both cognitive and motivational aspects as important factors in the expertise learning process, in contrast to traditional models which are "coldly cognitive" (Pintrich et al., 1993). Thus, it focuses on three components (i.e., knowledge, strategic processing, and interest) which are interrelated and play a role in this transition. In this paper, we focus on the first two components.

According to this model acquiring expertise has three stages (i.e., acclimation, competence, and proficiency/expertise). The most critical differences between experts and novices is the development of a broad and deep knowledge base, a shift in the kind of strategies used, from surface-level to deep-processing, and the increase in individual interest which permits experts to maintain a high level of engagement over time. The concurrent development of these components allows experts to be actively engaged in problem finding, posing questions, and instituting investigations that (sometimes) push their domain boundaries.

In terms of expertise studies examining scientific thinking in particular, the research have gone through considerable changes since its earliest period where the focus was on testing circumscribed aspects of the scientific discovery process (e.g. Wason, 1968). Within the last two decades, researchers have used simulated discovery tasks in complex domains in order to track participants as they explore, test hypotheses via experimentation, and acquire new knowledge in the form of revised hypotheses (Schauble, 1996).

However, such empirical research has a number of problems when used to generalize about scientific reasoning, such as the fact that scientific research is a collaborative enterprise taking months or even years to complete (Dunbar, 1995). Thus, researchers in the field of cognitive psychology (Dunbar 1995; Nersessian et al., 2003) and education (Bond-Robinson and Stucky, 2005; Feldman, 2008; LaPidus, 1997) have entered the laboratory to study scientists as they pursue research in real time.

Unfortunately, almost all of this research is weighted towards laboratory-based, experimental sciences; in fact, the only studies centered on field-based sciences are by Bowen and Roth (2007) in their examination of field ecologists. Thus, in this study we decided to focus on a field-based science with a strong *historical* component in order to better understand how expertise is learned in such fields.

As opposed to experimental-based sciences, such as chemistry which pursue experiments on natural phenomena under controlled laboratory conditions, the goal of *historical sciences* such as geology is to reconstruct past phenomena based on (mostly) un-manipulated evidence gathered as traces from the field. In fact, historical-based scientific methodologies were specifically developed to cope with problems that could not be solved experimentally. This has required the development of a whole set of methodological tools (both theoretical and practical) that help historical-based scientists overcome the constraints of their field-based evidence (Dodick and Argamon, 2009).. It has also affects the way graduate students are trained in the expert methodologies that historical scientists use to solve research problems, as we will show here.

## Methods

### The Research Sample

The research field of the team we investigated is *Integrative archeology*; a relatively new scientific domain, its goal is to use the (chemical and physical) properties of the materials accumulating at an archaeological site to better understand the humans who inhabited that site. Such materials include minerals, and organically produced materials such as bones, teeth, shells and plant remains.

The team's advisor (senior Professor) is trained in geology, biology and chemistry and has published extensively in the scientific literature. His team, at the time of this study consisted of two interacting groups:

expert scientists (2 junior colleagues), and novice scientists (1 M.Sc. and 3 Ph.D. students). Additionally, technicians, post-docs and summer students joined the team during summer fieldwork sessions.

As he advocates an interdisciplinary approach, the advisor accepts graduate students from most scientific fields; this approach is reflected in the team's background. "E" and "R", the two junior colleagues are trained in physics and archeozoology, respectively. Amongst the Ph.D students, "L" completed an M.Sc in Molecular biology; "D" has an M.Sc in Archeology and "A" in Materials Science. In this study we focus our analysis on the advisor and his 3 Ph.D. students.

## Data Collection and Analysis

The team pursues fieldwork at least once a year in order to collect samples for their analyses. To collect our data, we joined the groups' excavations (which typically lasted three-weeks during the summer) during a 4.5-year research period (2004-2009). We videotaped nearly 50 hours of interactions among research team members as they gathered and analyzed data in the field. Extensive field notes complement the video data. We also recorded the groups' seminars where members presented their works.

Semi-structured interviews were held with the team members after each field season. The interviews informed us of the researchers' individual feelings and understandings about various aspects of their research. It was also used to refine and validate our understanding of the field observations. All of the interviews were recorded and transcribed.

The interviews provided us with rich amounts of data touching upon a considerable number of themes which were analyzed using Shkedi's (2004) constructivist (ethnographic) method of qualitative research. In the first stage of this analysis, primary categories were developed following a first reading of the interviews allowing an initial coding. Careful attention to these categories produced a focused emergent framework, and the data was then recoded according to that framework. Our analysis concentrated on the major research challenges that the group members faced, as well as the strategies they employed to cope with these challenges.

A second analysis was done on the interviews and field observations using the questions that the advisor and students generated spontaneously about their research. Questions are identified as "an interrogative sentence or a declarative sentence with an embedded interrogative" (J. Dillon pers. comm.).

Two coding schemes were used in this analysis: (1) An emergent scheme (Table 1) which reflects both the questions connected to the research goals of the team, as well as to the strategies the advisor mentioned as promoting research. (2) A deductive scheme (Table 2) based on Dillon's (1984) classification of (scientific) research questions. Inter-rater reliability of these two schemes averaged 90%.

Table 1: Emergent categories of the questions asked by the team members.

| Category and Definition | Example |
|---|---|
| **Validity** questions are concerned with the extent to which the results of a research study can be generalized to situations beyond those involved in the study (*External validity*) **or** about the extent to which extraneous variables have been controlled by the researcher, so that any observed effects can be attributed solely to the "treatment" variable (*Internal Validity*). | What happened to the organic material since its initial deposition? (This question concerns whether the material changes were caused by chemical degradation.) |
| **Methodological** questions are concerned with procedures needed to solve a specific scientific problem. | How do we date plaster? |
| **Human Behavior** questions are concerned with attributes connected to human artifacts (e.g. Ceramic tools) or activities (e.g. Ways of cooking). | What were these ceramic vessels used for? |
| **Materials** questions are concerned with attributes of materials (e.g. color, IR spectra measurements). | What will be the ν-*Ratio* of the samples? |
| **Space and Time** questions are concerned with the spatial and temporal relationship between items such as Loci (areas), Layers, and Samples. | Which Archeological layer is above which? |

Table 2: Deductive scheme based on Dillon's (1984) categories for classifying scientific research questions

| Category and Definition | Example |
|---|---|
| **Property** questions are concerned with the properties of a specific "thing" (such as its shape, color, location). | What are the organic materials that were in this ceramic vessel? |
| **Comparison** questions deal with the comparative attributes between two "things". | What is the difference between the Calcite which was heated and the one that wasn't? |
| **Contingency** questions deal with the relationship (relational, correlative, or causal) between two "things". | Are the Copper rich sediments in situ or are they a product of post-depositional process of melting in water and going through the sediments? |

## Results and Discussion

### The Team's Research Process

Even though fieldwork happens only for a short period each year, it is critical to the team's research process (Figure 1). In the first research stage, physical and chemical traces of human activity are collected, analyzed in the (portable) lab and than reapplied to the field in order to locate a relevant research problem. In the second research stage, specific traces are collected in selected locations, followed (sometimes) by simulations in the lab and reapplication of the results to the field in order to solve the research problem. Both stages require at least two iterations, because lab analyses often require the team returning to the field and re-sampling a specific area in order to clarify and validate results emerging from the initial analysis.

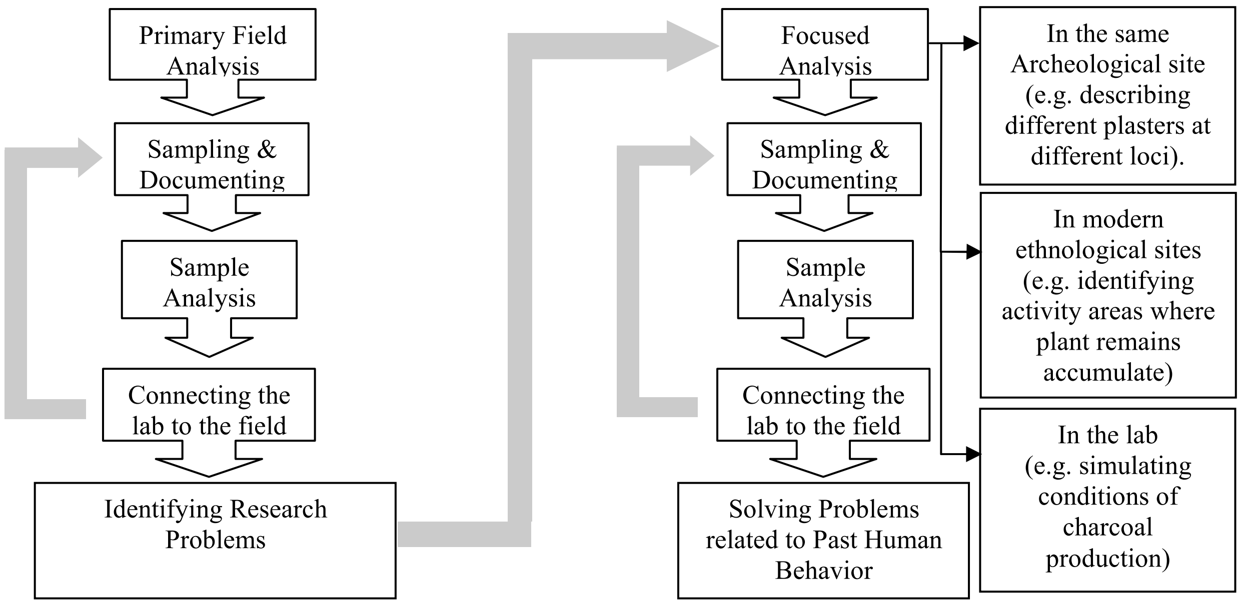



Figure 1: A schematic model of the team's research process

An example of a typical research problem was based on the observation that at certain site locations many of the sediments analyzed by IR-spectroscopy seemed to have been chemically altered by exposure to high temperatures. Solving this issue involved fully characterizing the phenomenon at different geographical and stratigraphical locations of the site, as well as modeling the effect of heat on sediments under both lab (using ovens) and field conditions (using natural fire); such simulations are intrinsic to the methodology of historical sciences. The aim of this problem-solving process was to create a tool enabling the researchers to reconstruct human behaviors from field materials, both at this site, and possibly at other sites as well.

## Assessing the Development of Expertise amongst the Students

Our data indicates that the advisor strongly stresses (24 times in a total of five interviews) the importance of problem identification as possibly the most critical element of field research and the most difficult one to teach (Dodick and Flash-Givili, 2008). Typical is his comment after the 2004 field season.

> This is "the state of the art" or "the name of the game" in science! It's not to collect observations, to summarize the details and to publish it; however, in these complex situations, without a direction, to identify the “thread” that will lead to something significant. If this would be something that I could teach in a lecture or a class...it would have been useful, than I think I don't need to teach further; the rest is technique."

We frequently noticed the advisor teaching this principle to his students. A typical example is his feedback (from field season 2006) to two of his students who immediately tried to solve the questions connected to a specific section even before mapping it and identifying the questions it elicited. "This is the wrong picture! In other words, even if we where there [at the problem solving stage] it's not for... I mean, if it's a serious problem we will do it, we will solve it, we will ask these questions, but maybe it's not even stratified..". Later, when asked to reflect on this event, the advisor said: "The big message is that you have to use the tools in order to do it in iterations, every time having a smarter question. Not in an automatic way." Thus, 'listening' to the questions presented by the field is a leading research strategy for him.

These observations led us to hypothesis that if we analyzed the questions posed by the students we would be able to better understand the progress they made in their research. Hence, we characterized the frequencies of questions asked by the three Ph.D. students' in the interviews (Table 3). Note that in 2006, the researchers surveyed a second site, in addition to the main focus site with which they began their research. It is important to take into consideration though, that this quantitative account is just a raw indicator of trends. This is not meant to be a fully quantitative study because of the small group size.

Table 3: Changes in the types of questions posed by the 3 Ph.D. students after each field season based on our emergent categories (presented as percentages of the total)

| | | **Interview 1** | **Interview 2** | **Interview 3** | **Interview 4** |
|---|---|---|---|---|---|
| | | 2004-2nd year* | 2005-3rd year | 2006-3rd year | 2007-4th year |
| Student A (PhD) | | N=13 | N=24 | N=17 | N=15 |
| | Materials | 38% | 17% | 40% | 40% |
| | Hum. Behavior | 62% | 29% | 12% | 27% |
| | Space & Time | 0% | 37% | 24% | 20% |
| | Validity | 0% | 0% | 18% | 0% |
| | Methodology | 0% | 17% | 6% | 13% |
| | | 2006 - 1st year* | 2007-2nd year* | 2008-3rd year | 2009-4th year |
| Student L (PhD) | | N=15 | N=20 | N=19 | N=12 |
| | Materials | 20% | 35% | 52% | 50% |
| | Hum. Behavior | 13% | 5% | 5% | 0% |
| | Space & Time | 53% | 20% | 0% | 8% |
| | Validity | 7% | 10% | 16% | 0% |
| | Methodology | 7% | 30% | 27% | 26% |
| | | 2005-4th year | 2006-4th year | | |
| Student D (PhD) | | N=30 | N=23 | NA (Ph.D. completed) | NA (Ph.D. completed) |
| | Materials | 13% | 17% | | |
| | Hum. Behavior | 30% | 17% | | |
| | Space & Time | 33% | 43% | | |
| | Validity | 17% | 17% | | |
| | Methodology | 7% | 6% | | |

Looking at the data, a number of important trends are seen. In the case of “A” there was a strong tendency at the beginning of her work to focus on Human (cultural) behavior. This is well reflected in the following quote from an interview after field season 2004 (the second year of her Ph.D.):

> I want to work on things: ceramics, tools, kilns, installations, technology, to understand technology, to see what happens to it, which materials did they use? Why did they use them? Where did they take it from? Those kinds of things…

These types of questions, as she noted, are not encouraged by the advisor as he prefers asking questions which will enable his team to create a universal tool, or research strategy which can be applied to many sites. However, even though she recognized his intentions, she still insisted on focusing on human behavior questions. This may have been due to her misconception that her immediate task, as an archeological problem-solver was to reconstruct human behaviors of the past. This also correlates with our field observations which showed that at the beginning of her Ph.D. this student worked very closely with the traditional archeologists, and this too may have swayed her thinking towards human behavioral problems. In fact, her advisor mentioned that at the beginning of her work, he needed to "separate" "A" from the archeologists".

However, the percentage of this question type tended to tail off through the third year of her Ph.D., as she began to understand the importance of analyzing the properties of field materials both in terms of content and their spatial-temporal relationships. This is clearly seen in the following quote from field season 2007, which was the fourth and final year of her Ph.D. research.

> Today I don't need someone to explain to me to know that this [layer] cuts this and this cuts the other, that this is early and this is late. This might sound trivial, but it wasn't at all like that for me at the beginning. I also have a little bit more understanding of the processes, of how they occur. Which kind of by-product you expect from a pile of dung which had disintegrated, or from metal that stayed in the ground. What to expect if I have a collection of metals? How the does the sediment around it look? All these little contexts are much clearer to me today.

In contrast to "A", "L" asked a smaller frequency of human behavior questions even at the beginning of his Ph.D. and the frequency of such questions rapidly declined through his research. Concomitantly, we see an increase in the percentages of materials, validity and methodological questions; this increase connects well to his interest in methodological problems, in which the ultimate goal is to create a tool that might be used to better reconstruct the temporal background of a site. As he noted in the first year of his Ph.D. studies: "If we could invent a scientific physical method that will really prove that one ceramic artifact is present before the other [...] it's a very nice idea to do". Indeed, "L's" research focused on validating a new tool for dating plaster compounds. Interestingly, at the end of 2008, his research focus shifted toward characterizing the chemical properties of plaster with no connection to developing a new tool. This change was forced upon him by circumstances in the field, as he encountered some major difficulties in the dating project, while at the same time he started identifying extraordinary examples of plaster in a specific site.

In essence, the shift in "A" and "L's" questions represent the influence of the advisor's research philosophy which emphasizes the importance of starting a project with a field-based material analysis before jumping to questions of human behavior or trying to invent a new tool. However, at least in the beginning of their studies, his students had a tendency to pose broad questions (i.e. what did people do here?) and immediately tried to solve them. Moreover, they were inclined towards using "actualistic" logic to answer their questions even before collecting data in the field. (Actualistic logic is used to reconstruct the past based on knowledge of present conditions. It is a widely used in all historical sciences). Indeed, the advisor often had to warn them "not to use their logic to come to a conclusion". Instead, he emphasized the importance of "using logic to ask a question" and "to try and get the data to support it".

Finally, "D" like her colleagues shows a decrease in human behavior questions, but unlike "A" and "L" she shows almost no change in materials-based questions. This may be due to her archeological background, which made it hard for her to look at the field from a materials perspective. Indeed, she admitted that although she had learned much in her Ph.D. she still needed to fill in considerable gaps in her understanding of chemistry. This connects well to the way her advisor characterized her progress towards her Ph.D., which was in the direction of becoming a "lab specialist"; this, contrasts with his research strategy which integrates both field and lab findings.

Another research element emphasized by the advisor (2006 interview) is connected to understanding the spatial and temporal context of the field, as he noted: "Excavation for everybody is the bread and butter of Archeology. I excavate, I can decipher the stratigraphy and can complete the picture. If you are doing mistakes in deciphering the stratigraphy than it's a big mess, that nobody can work out". In Table 3 we see that "A" asked no spatial-temporal questions at the beginnings of her Ph.D. because as she admitted she "didn't even understand where to start observing". This is supported by the lack of any stratigraphic diagrams in her field notebook at this time. It was only in the third year of studies that she began to ask relevant stratigraphic questions from her colleagues, and even began to draw field-based diagrams. In contrast with "A", "L", was highly focused on developing a good understanding of the spatial-temporal relationships of the field, and this is seen by the large frequency of questions he asked from the beginning of his studies. In fact, his advisor (interview, 2006) noted that he was “ready made for this type of work”. Finally, Student “D” was very focused on spatial-temporal relationships. This correlates well with her previous research experience in archeology.

We also used Dillon's (1984) scheme for classifying the questions; here, questions are classified to three levels: Properties (I), Comparisons (II), or Contingencies (III). The purposes of this teams’ research are identifying archeological materials and reconstructing human behavior; in contrast, the other types of questions (Validity, Methodology, and Space and Time) serve these purposes. For this reason, we decided to use Dillon’s (1984) scheme to focus in on materials and human behavior questions.

Our results show a very strong trend in (almost) every year, and every student towards level I Properties questions, both in the case of materials and human behavior questions. "A"'s Properties questions averaged 64% of her total materials-based questions across all the years we observed her, while her Properties questions devoted to human behavior averaged 71%. This situation repeats itself with "L" (77% and 100% respectively) and "D" (86% and 70% respectively). This supports the fact that the advisor is very careful in establishing the baseline properties of archeological materials before jumping to causal relationships (Level III questions). For example, one of his big research projects has been examining the material contents of an ancient (garbage) pit. However, much of the discussion connected to this work has been on defining the nature of “pits” in general. A similar issue arose around the definition of “metallurgic kilns” which was part of “A’s” research.

These results show that a good scientific research program does not necessarily have to concentrate on causal questions; establishing a phenomenon's properties is critical in of itself. This is especially important in field-based, historical sciences, where one doesn't control the evidence offered by the field; this forces the scientists to collect every important trace, while validating its properties prior to answering any causal question. In agreement with this observation, student "L" described the most important quality of this group's method as "Differentiating between [the phase of] documenting the data from its interpretation; first explain what was found, bring pictures, document everything, do everything properly and only after that stage, write in the conclusion what you think it is". This emphasis is also supported by the team’s publications, most of which have a strong descriptive character. The advisor hopes that his method will be applied to other archeological problems, providing his work with greater validity.

Finally, thematic analysis of the advisor’s interviews shows an abundance of strategies that he uses to identify research problems demonstrating that he is highly engaged in problem identification. Such strategies are either derived from the *disciplinary culture* (i.e. Identifying gaps in experts' knowledge) or the *data* itself (i.e. Following unexpected findings; Focusing on prominent data). However, at least at the beginning of their studies, most of the students’ strategies were not derived from *data* but rather from the *disciplinary culture* wherein they focused on problems previously identified by others; in other words they often expected their problem to be well defined by others, before collecting data. This is not surprising in that this is the first time that these students received an opportunity to fully develop a research program. Still, as they progress through their studies, the students adopt a more experiential based approach (i.e. focusing on data they know how to analyze, based upon 'what the field offers'). This shift demonstrates their transition towards greater expertise.

Our results correlate well with Feldman (2009) where he classified the highest level of graduate science research as creating "Knowledge Producers" who can, among other things, independently formulate research questions. Similarly, Alexander (1997) also defined the highest level of academic research (in general) as proficient experts who are actively engaged in problem finding. It also connects well with Van Gog et. al.’s (2005) research which showed that in contrast to novices, experts invest far more time representing a problem before solving it.

Nevertheless, our case study shows that such problem identification can be circumvented by the students’ misunderstanding of where their research truly begins, as they believed that their primary goal was solving human behavioral problems, when in fact they (first) needed to do a complete material analysis. Simply put, they attempted to by-pass a significant research stage. By misinterpreting their research goals they also sidetrack the process of formulating the questions that would ultimately guide their studies by immediately jumping into a problem solving mode. This is reflected by how their questions change over time.

These results emphasize that the first goal of inquiry learning, i.e., "identifying questions that can be answered through scientific investigation" (NRC, 1996) is a complex skill that novices need to develop before they become scientific experts. Thus, even before dealing with the second goal of inquiry learning, "design and conduct a scientific investigation", amongst which controlling variables is listed as a major skill, we need to invest more time in supporting students with experiences that will develop their ability to formulate effective research questions (Kuhn and Dean, 2005); we certainly cannot expect them to develop this skill by themselves.

Obviously this is a small, focused case study so more research needs to be done to provide increased validity to our results. Thus, we have started a comparative study analyzing another field based science – ecology. This will permit a fuller description of some of the in-depth problems that students experience while learning to become expert field scientists.

## References


Alexander, P.A., (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32(8), 10–14

American Association for the advancement of Science (1967). *Science – A process approach*. New York: Xerox Division, Ginn and Company.

Bond-Robinson, J., Stucky, A.P. (2005). Grounding scientific inquiry and knowledge in situated cognition. In: *Proceedings of the 27th Annual cognitive science society* (pp. 21–24), Stresa, Italy.

Bowen, G.M., Roth, M.W. (2007). The practice of field ecology: insights for science education. *Research in Science Education,* 37, 171–187.

Brill, G., & Yarden, A. (2003). Learning biology through research papers: A stimulus for question-asking by high-school students. *Cell Biology Education*, 2, 266– 274.

Chi, M. T. H., Glaser, R., and Farr, M. J. (1988). *The Nature of Expertise*. Erlbaum, Hillsdale, NJ.

Chin, C. (2004). Students' questions: Fostering a culture of inquisitiveness in science classrooms. *School Science Review*, 86(314), 107– 112.

DeGroot, A. D. (1946). *Het Denken van den Schaker. (Thinking Processes in Chess Players.)* Noord Holland, Den Haag, The Netherlands.

Dodick, J., & Argamon, S. (2009). Understanding scientific methodology in the Historical and Experimental sciences via language analysis. *Science and Education*, 18 (8), 985-1004.

Dodick, J., and Flash-Gvili, I. (2008). Challenges to graduate student research in the historical based sciences. Proceedings of the National Association of Research in Science Teaching Meeting, Baltimore, MD.

Dunbar, K. (1995). How scientists really reason: scientific reasoning in real-world laboratories. In: Sternberg RJ, Davidson J (eds), *Mechanisms of insight* (pp. 365–395). Cambridge MA: MIT Press,.

Ericsson, K.A., Kramp, R.Th., & Tesch-Romer, C. (1993) The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.

Feldman, A., Divoll, K., Rogan-Klyve, A. (2009). Research Education of New Scientists: Implications for Science Teacher Education. Journal of Research In Science Teaching, 46 (4), 442–459.

Gobet, F., & Simon, H.A. (1996) Templates in chess memory: A mechanism for recalling several boards, *Cognitive Psychology*, 31, 1–40.

Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, 42, 791-806.

Kuhn, D., & Dean, D., Jr. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16, 866–870.

LaPidus, J.B. (1997). *Doctoral education: Preparing for the future*. Washington, DC: Council of Graduate Schools.

Marbach-Ad, G., and Sokolove, P.G. (2000). Can undergraduate biology students learn to ask higher questions? *Journal of Research in Science Teaching,* 37, 854–870.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council (2000). *Inquiry and the national science education standards*. Washington, DC: National Academy Press.

Nersessian, N.J., Newstetter, W.C., Kurz-Milcke, E., Davies, J. (2003). A mixed-method approach to studying distributed cognition in evolving environments. In: *Proceedings of the international conference on learning sciences* (pp. 307–314).

Pintrich, P. R., Marx, R., & Boyle, R. (1993). Beyond "cold" conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167–199.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. Developmental Psychology, 32, 102–119.

Shkedi, A. (2004). Words of Meaning: Qualitative Research-Theory and Practice. University of Tel Aviv Press: Tel Aviv.

Sternberg, R.J. (2003). What is an "expert student?" *Educational Researcher*, 32(8), 5-9

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, 19, 205-221.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology,* 20, 273-281.


**Acknowledgements:** This research was supported by Israeli Science Foundation Grant 731/06, the Levi Eshkol Institute for Social Sciences Research Award of the Hebrew University of Jerusalem and the Municipality of Jerusalem Student Research Scholarship.

# Teachers Collaborating with Wiki: The Impact of Professional Status, Language, and Age

Yael Poyas, Oranim College of Education, 36006, Israel, yael_p@staff.oranim.ac.il

**Abstract**: Research indicated that Wiki-aided teaching has many advantages if supported by appropriate pedagogy. The present study examined the effect of this environment on the learning process during an M.Ed literature course for a multidisciplinary group of Jewish and Arab teachers in Israel. The findings were derived from the Wiki data, from learners' written and oral feedback regarding the experience, and from the lecturer's reflective diary. Although the learners' and lecturer's satisfaction was high, since the Wiki framework contributed to the learners' interest, involvement, and depth of their investigation, yet self-confidence in language use, cultural learning habits, age and professional status affected the learners' performance.

This article is about the use of the Wiki technology as an alternative platform for the construction of a shared interpretive space (Sumara, 2002) for the study of literature in the context of an M.Ed program for experienced teachers. It examines the data from the point of view of a teacher educator, investigating the experience of her teaching with the purpose of both improving her work and of contributing to the accumulating knowledge regarding the use of Wiki in professional development programs for teachers.

## Theoretical Background

### The Wiki Environment in Educational Settings

The Wiki platform is an environment enabling cooperative and collaborative learning, each participant being permitted to write, add, edit and alter any text written within its framework (Leuf & Cunningham, 2001). In this sense, Wiki's features create a "low risk" editing environment (Wang & Beasley, 2008), providing a multimodal and hypertextual writing platform, transparent to all readers. Studies examining Wiki-aided teaching found that this environment has many advantages. It fosters constructivist learning, as well as collaboration and interaction among the students themselves and between each student and the teacher. The knowledge constructed by students is made available to all the participants in the course for study and evaluation (Morgan, 2004; Watson, Boudreau, York, Greiner & Wynn, 2008). Moreover, it appears that the level of investigation by students using the Wiki platform improves, due to their being exposed to appraisal by their colleagues and others (Ravid, 2006). Recently researchers recommended the incorporation of Wiki as a platform for creating collaborative updated and course-oriented textbooks, in order to empower learners (Ravid, Kalman & Rafaeli, 2008). As Ruth and Houghton (2009) claim, Wiki is more than a tool, it is a way of learning.

Satisfaction with learning in the Wiki environment was found to be high, when supported by appropriate pedagogy (Ben-Zvi, 2007; Meshar-Tal & Tal-el-Hasid, 2006; Morgan, 2004). Studies dealing with learning language and literature with the help of Wiki show that this environment promotes high-level reading comprehension, a diversity of interpretations, and dialog among interpreters with various cultural backgrounds (Désilets & Paquet, 2005; Faranbaugh, 2007). It also increases writers' awareness of the effectiveness of critique in improving their written products, their meticulousness regarding spelling and syntax, and their sensitivity to different context-dependent modes of writing (Wheeler & Wheeler, 2009).

Researchers reveal that the success of educational Wiki-based learning projects or their failure depends on many factors; indeed, the same researchers may sometimes report success and at other times describe difficulties and discontent (Wang & Beasley, 2008). It was also reported that the learners' attitude towards Wiki and their time investment in posting varies from one learner to another (Ravid, Kalman and Rafaeli, 2008; Robertson, 2008). Studies also showed that the specific discipline and its culture of study affect the success in Wiki-aided learning (Rick & Guzdial, 2006).

### Culture and learning via interactive educational technologies

Culture influences its members' ways of thinking and interpreting (Branch, 1997; Matsumoto, 1996). When teaching with the aid of technological tools two cultural factors should be noted: effects of cultural attitudes concerning interpersonal communication, and the language and ways of coding and decoding knowledge (Wild, 1999), as well as the effects of the learners' epistemological and philosophical perceptions concerning teaching and learning.

*Effects of cultural attitudes* - Researchers proposed a variety of models to explain the differences between cultures, for instance between individualist and collectivist cultures, cultures with high-level context-dependent communication as opposed to cultures with low-level context-dependent communication, as well as cultures capable of accepting ambiguity and those avoiding it (Gunawardena, Wilson & Nolla, 2003; Hall, 1966, 1976; Hofstede, 1986; Matsumoto, 1996). Different cultural codes, culture-dependent information processing and expression, as well as culture-dependent networks of relationships and norms of interpersonal communication, may cause participants various problems while studying in multicultural groups in social technological interactive environments such as Wiki (Gunawardena, Wilson, & Nolla, 2001; Hall, 2006). There are cultures (like western cultures) in which language is perceived as a tool conveying precise meaning, and other cultures (like the Arab culture), where it is perceived as a system of linguistic forms, arousing feelings and visual images. Thus, while one culture pursues functional, precise, explicit and unambiguous use of language, another prefers affective language, implicit and rich in imagery (Dwairy, 2006; Zaharna, 1995). For instance, studies showed that students from cultures avoiding ambiguity find online learning environments more frustrating than students from cultures accepting ambiguity (Downey, Wentling, Wentling & Wadsworth, 2005), and that students writing in a foreign or second language participated less and were less confident in their contributions (Yilditz & Bichelmeyer, 2003).

*Effects of learners' perceptions concerning teaching and learning* - Studies report that while developing their pages on Wiki, many learners find it difficult to cope with the task of knowledge construction owing to the norms they had become used to, with the teacher not only providing them with the required knowledge, but also determining the framework for its organization and presentation (Farabaugh, 2007; Wang & Beasley, 2008). Such difficulties become acute when students, partaking in a collaborative interactive multicultural learning group, are accustomed to traditional spoon-feeding ways of teaching and learning; they encounter tremendous obstacles in attempting to adapt to the self-directed environment (Alsunbul, 2001). The characteristics of academic culture, which honors ownership of knowledge and copyright laws, were also found to affect activity in the Wiki environment; they minimize learners' willingness to allow peers to edit their Wiki pages and their own willingness to evaluate other students' work (Lindsey, 2006; Wang & Beasley, 2008).

*Educational technology and teachers' development*- Research on teacher education and educational technology including Wiki is growing rapidly. The teachers' attitude towards the benefits of technology for everyday school instruction is a crucial factor affecting technology implementation (Becker, 2000; Cuban, 2001). If the experience of learning with Wiki undergoes reflective processing and is integrated in the teachers' curricular thinking, it may be translated into effective use in the planning of teaching and in sharing experiences with colleagues (Darling-Hammond, Banks, Zumwalt, Gomez, Gamoran Sherin, Griesdorn & Finn, 2005). A positive experience of the use of the Wiki technology in teacher education programs will hopefully encourage teachers to subsequently use Wiki in their teaching.

The current study focuses on the impact of learning with Wiki on the process of the learners' writing and editing of their contributions as a result of a collaborative activity, within the framework of in-service M.Ed studies at a college of education in northern Israel.

## The Study

### Course Context

The M.Ed program of multidisciplinary instruction is intended for experienced teachers from Arab and Jewish schools from the northern periphery of Israel. The Arab and Jewish sectors in Israel have separate elementary and secondary educational systems, each one of them having its own language of instruction (i.e., Arabic, Hebrew). Only in higher education are Jews and Arabs engaged in multicultural encounters, sometimes in very conflictual political contexts. Studies examining the integration of Arab students in colleges and universities where the language of instruction is Hebrew, found that Arab students had to cope not only with high level and academic Hebrew, but also with reading material in English, teaching methods demanding more independent study, and more free and open relationships between students and between them and their teachers (Al-Haj, 1996; Peleg & Raslan, 2003).

*Course Participants* - 19 participants took part in the course under study, 9 Arab teachers and 10 Jewish teachers. Most teachers were women (typical of the state of the profession in Israel). Their ages ranged from 26 to 56 years, while half of them were between 41 and 55 years old. The teachers' professional experience ranged from three to dozens of years. Twelve of the teachers were teaching in primary schools and the rest in secondary schools. Only seven of them had studied literature for their B.A. degree. None of them had previously read the novel studied, nor had any experience of study in the Wiki environment. The affiliation of the group's members to Arab or Jewish society, their different mother tongue (i.e., Arabic or Hebrew), were the salient cultural factors distinguishing between them. As for other factors, such as age and professional identity, the participants had much in common.

*Aim and Content of the Course* - The aim of the course under study (14 weeks) was to engage the learners in an encounter with world literature (Cai & Sims Bishop, 1990), and to build up a group database about the novel *The Day Lasts more than a Hundred Years* by Chinghiz Aitmatov, sharing historical, geographical, cultural, and literary knowledge, as well as individual and group interpretations. The assignment was to write an entry about a broad topic in the novel, relating it to entries written by other colleagues, complementing or expanding them. Writing on the Wiki platform evolved alongside 10 weekly face-to-face lessons, lasting 90 minutes each. (In the other four weeks there were no F2F meetings.) Blended model of teaching & learning was preferred to reduce the difficulties of adapting to Wiki-based writing, and to enable whole group F2F discussions of the novel and the written products. The learners were also asked to respond to their colleagues' entries on the 'discussion board'. The writing process continued until a month after the end of the course sessions, altogether about five months. No specific guidelines for content, style and organization of the entries were provided, in order to give the learners freedom with respect to their style and interpretations.

### Research questions

This paper focuses on the development of the written products on the group's Wiki database.

1. What was the pace of the entry into activity in the Wiki environment and which factors affected it?
2. What steps did the learners take and which strategies did they employ in their writing when given free rein to carry out literary interpretive tasks in the Wiki environment?

### Data Collection and Analysis

The data were derived from three main sources:

*Data provided by the Wiki platform:* (a) The entries written by the teachers; (b) The remarks on each entry, written on the discussion board; (c) Data provided by the history of each entry regarding the number of posts, changes and editing operations made.

*Data derived from the teachers' reactions:* (a) The teachers' remarks in the course of the lessons, noted down in the lecturer's diary; (b) Feedback reports written by each teacher about his/her experiences; (c) The transcript of the discussion that took place during the summing up of the Wiki-based experience.

*Data from the lecturer's diary:* The lecturer's own notes about (a) teachers' remarks, (b) her conversations with them, as well as (c) the phenomena and problems identified by her during the course.

The Wiki pages' history boards were analyzed in order to learn about the teachers' pace of writing and editing throughout the course. The written entries were surveyed for the number of linkages made by the teachers at three points of time – during the first version, the final one and an in-between version. Utilization of linkages to their peers' entries (and/or other Internet sites) implied teachers' adjustment to some of the norms of writing in hypertextual environments as well as their familiarity with peers' products and contributions. The discussion boards were examined to identify the contributors and the topics discussed. Learners' responses, collected throughout the course and derived from their feedback pages and the final discussion, were categorized according themes concerning participants' (a) experiences of success, (b) apprehensions, (c) difficulties, and (d) ways of coping with these difficulties. Learners' responses were used to investigate the considerations they had in mind and to explain phenomena, related to features characteristic of the Wiki platform. The lecturer's diary assisted her in recalling her reasoning regarding the learners' difficulties, the phenomena mentioned in their' statements, and the evolvement of the writing throughout the course.

## Findings and Discussion

On the whole, the teachers displayed enthusiasm and interest in the Wiki assignment, and gradually developed a Hebrew database about the novel. The teachers' feedback at the end of the course was positive, and all of them emphasized that Wiki had increased their involvement in the course. However, a closer examination of their strategies and activity called attention to some factors that may have impeded involvement and performance, and they should be addressed.

## The Pace of Entering into the Activity

There were differences, related to teachers' specific cultures and languages, in the pace of their work and their readiness to write the opening paragraph of the entry, as can be seen in the following Table (see Table 1).

Table 1: Points in Time of Students' First Entrance into the Wiki Environment

Weeks of the Semester

| 2 | | 3 | | 4-5 | | 6-9 | | 10-14 | | Later | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Arabs | Jews | Arabs | Jews | Arabs | Jews | Arabs | Jews | Arabs | Jews | Arabs | Jews |
| - | 1 | - | 3 | - | - | 2 | 4 | 6 | 2 | 1 | - |

According to the participants' explanations, the slow entry into writing on the website was due to four main reasons: (a) less confidence in their knowledge of the language; (b) the felt need to become well acquainted with the text; (c); their habits of study, and (d) their attitude towards technology.

*Knowledge of the language* was mentioned repeatedly by four of the Arab teachers, their fear of writing in a faulty style, of making linguistic and spelling mistakes for all to see. When asked, the Arab teachers who were actually teaching Hebrew (6 out of 9 teachers) stated clearly they were more fluent and proficient in Hebrew academic writing than in their own language; however, when it came to exposing in public their writing in progress alongside that of their Jewish colleagues, they became less willing to participate in the activity.

*Knowledge of the novel*: Native Hebrew speakers read the novel fluently and did not wait for group discussions in order to continue reading, while eight out of the nine Arabic native speakers read the novel chapter by chapter, relying on classroom discussions to clarify the novel's content and the issues it raised. As a result they constructed their initial impression of the novel at a later stage of the course and started writing their entries later than most of their Jewish colleagues.

*Habits of study:* Three (33%) Arab teachers expressed frustration regarding the complexity of the task, which called for their own interpretation of a phenomenon in the novel or outside it, without any specific guidance regarding content or structure. Their explanations exposed their expectations that the lecturer would provide them with clear instructions, similar to those usually provided in an academic course regarding the final paper. The case of these teachers' slower and more cautious entry into the activity and the crutches they needed drew attention to the difficulty, caused by open assignments in the transparent Wiki environment as experienced by those who are not confident enough to cope with multidirectional and unstructured tasks. The Wiki technology transparency leaves no private space to those needing more time and guidance; rather it "exposes" them to the entire community. When the community of learners consists of self-aware experienced teachers, this exposure may mentally block some of them.

*Attitude towards technology*: Some older teachers who hesitated in their first steps explained their difficulties in their reluctance to use technology for writing.

According to the participants' responses, it appears that their mother tongue, age as well as learning habits were important factors, affecting the pace of entry into the writing activity.

## Steps and Strategies in Writing

*Using the word processor:* About a third of all participants admitted that they wrote their thought first on the word processor, placing them on the website only when they were entirely satisfied with what they had written. This phenomenon may also explain why many entries began to appear on the website only in the last third of the semester. Seven teachers, Arab and Jewish, mentioned they were incapable of showing their work in progress to others. Seven others began writing only after composing several paragraphs on the word processor, but from this point on continued to use the Wiki, without resorting to the word processor again. Only four Jewish and one Arab teacher presented their work in progress throughout the course. During the feedback discussion at the end of the course all the teachers agreed with one of their colleagues' assertion that it was very strange and unusual to skip the private, intimate and reassuring step of working on the draft papers, and that it needed courage to expose partial products in public.

Most participants (70%) reported that their apprehension regarding their colleagues' reactions affected their writing: "I am a teacher; I evaluate and give grades to my students. Am I supposed to let others see a slipshod piece of my work? That's totally out of the question. It took me a relatively long time to accept that it is possible to post on Wiki ideas not properly thought through. You [the lecturer] must understand the radical change we had to undergo to participate in Wiki". Another student remarked: “Writing is a very private matter for me. I can’t imagine exposing to others the process of the consolidation of my ideas”. Such remarks highlight the radical change in attitude required for work in the Wiki environment regarding the long-established habits that experienced teachers acquire in the course of their studies and professional development.

*Duration of writing, editing and revising:* The students not only started their task at different points in time, but also varied in the number of times they posted and edited their entries and in the duration of their Wiki-based writing periods, as is evident on the following Table (see Table 2):

Table 2: Length of time learners worked on Wiki and word processor assistance

| Duration | Arab Learners | Jewish Learners | Assistance of word processor |
|---|---|---|---|
| 4.5 Months | - | 4 | No assistance |
| 4 months | 2 | 4 | Partial assistance |
| 3 month | 3 | 1 | -"- |
| 1 month | 2 | - | Full utilization, with final editing work on Wiki |
| 1 Week | 1 | - | Full utilization |
| 3 days | - | 1 | Full utilization |

Teachers who wrote directly on the website edited their entries fifteen to twenty times in the course of the semester, worked on it for long periods and expressed explicitly their great involvement in the novel and their enjoyment of the course. Others posted or edited their posts only 3 to 7 times.

*Writing in a hyper-textual environment*: At first the entries resembled the usual type of academic writing, the students almost totally ignoring the hypertextual features of the Wiki environment. The awareness of the need to adapt their writing to a hyper-textual environment developed slowly. An overall survey of the entries' development revealed that it took seven or eight weeks before the participants realized what a hypertextual environment could offer. It occurred only after (a) the entries had grown to several paragraphs each; (b) participants were explicitly instructed to read their colleagues’ entries (c) devoting time to F2F review of several entries. At the same time the paragraphs in their own entries became denser and more in line with the definition of their topic. However, only small number of links was made by learners to external sites and the use of internal links was moderate (see Tables 3 and 4).

Table 3: Number of links to peers' entries related to specific points in time during the course

| | | No. of Links to peers' entries: 0 | 1-3 | 4-7 | 8-12 | 13-14 |
|---|---|---|---|---|---|---|
| No. of learners | Initial version | 13 | 3 | 1 | 1 | - |
| | In-between version | 2 | 5 | 10 | 2 | - |
| | Final version | - | 4 | 11 | 3 | 1 |

Table 4: Number of external links related to specific points in time during the course

| | | No. of external links: 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| No. of learners | Initial version | 18 | - | - | 1 | - | - |
| | In between version | 11 | 3 | 3 | 1 | 1 | - |
| | Final version | 8 | 5 | 3 | 1 | - | 2 |

This phenomenon may be related to traditional academic writing habits and routines, and to difficulties to adjust to interactive hypertextual norms of writing. Reflections of six of the Jewish participants support this assumption.

*Contribution of exposure to colleagues’ entries and to discussion boards:* All participants mentioned in their feedback that reading their colleagues’ entries led them to consolidate their own and concentrate on the topic as defined. It enriched them with ideas and offered them new directions of interpretation. They had also found reading the conversations on the discussion boards between their colleagues and the lecturer very interesting. These exchanges provided them with new avenues of thought as well as guidance and direction.

*The written products as private property and the 'discussion boards' as private study rooms:* Almost all the participants were sensitive to their ownership of their entry and their colleagues' interference. They perceived the written text they produced to be the intellectual property of the writer. Seven of them even signed their names at the end of their entries. They argued that exposing the entries to public view permits their colleagues to read, but not to edit them. In this case, the teachers' comprehension of the principle of the sharing of knowledge and of collaborative editing was not translated into practice.

The discussion boards served mainly as a space for the lecturer's comments rather than a shared space for discussing content and ideas among group members. Writing their own evaluation of a colleague's work on his/her discussion board was perceived - as one of the learners put it - like "an invasion of someone's private study"; therefore they tended not to write feedback. The feedback session revealed that the participants discussed the written products with their colleagues and made critical comments, but used other channels, such as F2F talks, phone or e-mail, channels that keep feedback as a reciprocal act, taking place between two people, not as a learning space for all participants.

Only seven women (5 Jews and 2 Arabs) responded to 8 of their colleagues' entries, expressing their interest, suggesting more references, discussing the contents, and elaborating on connections between the entries. Only three of the teachers who received responses from others continued to converse with them. Only 10 responded to the lecturer's.

## Conclusions

The findings of this case study demonstrate that collaborative construction of a database about a literary work being studied on the Wiki platform enriches the learners' appreciation by providing them with diverse perspectives, and leads to a sharing of ideas and thoughts in the wake of their reading. The Wiki characteristics expose all the participants to a whole gamut of opinions, voiced throughout the course, while crossing cultural and social boundaries, owing to the possibility of reading the texts as they accumulate, and thus become acquainted with their colleagues' interpretations. Without the collaborative technological framework, many interpretations might not have been voiced.

A more cautious perusal of the results revealed that alongside the apparent success, certain limitations and problems must also be taken into account when teaching with Wiki in programs aimed at experienced teachers' professional development; the same is true in the case of multilingual and multicultural study groups.

The current study focused on the first experience on Wiki of a heterogeneous group of relatively experienced and older teachers, from different cultures and schools; it investigated the way they made use of the various devices available to them, to improve their written product. These learners had their own perceptions about a teacher's role and were liable to cling to their habits, and to act according to the rules and customs of the school in which they were working and the education system they were used to.

When such a group of experienced teachers, aware of their professional status, is to assimilate Wiki-based learning, it encounters obstacles, stemming from their habitual way of studying, their teaching cultures (Feiman-Nemser, & Floden, 1986), as well as their entrenched perceptions of the appropriate relations between teachers and students and their differing roles in the classroom. The findings reveal that the issue of ownership of knowledge prevented colleagues from becoming involved in the editing of the written text in progress. Moreover, the perception that evaluation is the teacher's right and also his/her obligation to the learners, but not the learners' to their colleagues, reduced the social-constructivist value of the Wiki technology.

The sense of fellowship among colleagues is not conducive to mutual criticism in the presence of the lecturer, responsible for giving grades (see also Brett, Forrester, & Fujita, 2009). In this sense, Wiki-based learning calls for alternative ways of assessment, different from the traditional academic one.

The study reinforced previous findings, which maintained that the ways of writing in the Wiki environment differ from customary academic writing and apparently more than one exposure to Wiki is necessary for assimilating the Wiki ways of creating a network of connections (i.e., an 'intertext'), enabled by this technology. Most of the learners remained inside the pages of the "group's book" they had written about the literary work. This phenomenon may be attributed to learners' limited experience with hypertextual non-linear environments due to their long established habits of learning (Dziuban, Moskal, & Hartman, 2005).

The study found that learners in the Wiki environment, not writing in their mother tongue, feel diffident; here this factor was salient and affected the ongoing process of group learning, since a part of the group, mainly the Arab learners, did not share with the others the development of their written product, but preferred to start writing later and base the text on previous attempts on the word processor. This resulted in the phenomenon that their late contribution was not discussed at the group's meetings, did not benefit from their colleagues' thoughtful comments and did not enrich other learners' products.

When the learners are teachers who have a high opinion of their ability to write and construct knowledge, the fear of exposure of their weaknesses may cause them misgivings about early posting on the website.

*Limitations of the study*: This study suffers from limitations characterizing 'messy' field conditions, since it is impossible to differentiate accurately among variables. Moreover, it tells a story of a single case study. Nevertheless, every case study of Wiki-assisted teaching may enrich our knowledge concerning the Wiki impact on learning in diverse contexts. Additional factors, such as the lecturer's way of teaching, the literary work chosen (its length, complexity and contents), as well as the academic backgrounds of each learner, call for further research in this field while using more modes of inquiry, such as in-depth interviews of the learners and their teacher.

## References


Aitmatov, G. (2006). *Ve Ha Yom Einenu Cale (The day lasts more than a hundred years)*. Translated from Russian by D. Marcon & A. Porat. Tel-Aviv: Am-Oved (Hebrew).

Al-Haj, M. (1996). *Jewish-Arab Encounter at the University of Haifa*. Haifa: University of Haifa.

Alsunbul, A. (2001). *Education in the Arab world at the outset of the third millenium*. Alexandria, Egypt: University Library.

Becker, H. J. (2000). Findings from the teaching, learning and computing survey: Is Larry Cuban right? *Revision of a paper written for the January, 2000 School Technology Leadership Conference of the Council of Chief State School Officers*, Washington, D.C. Retrieved May, 8, 2009, from: http://www.crito.uci.edu/tlc/findings/ccsso.pdf

Ben-Zvi, D. (2007). Using Wiki to promote collaborative learning in statistics education. *Technology Innovations in Statistics Education, 1*(1), Article 4.
Retrieved April, 30, 2009, from: http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art4

Branch, R. M. (1997). Educational technology frameworks that facilitate culturally pluralistic instruction. *Educational Technology, 37*(2), 38-41.

Brett, C., Forrester, B. C., & Fujita, N. (2009). Online learning journals as an instructional and self-assessment tool for epistemological growth. *Canadian Journal of Learning and Technology, 35*(1). Retrieved 2 February, 2010 from:
http://www.cjlt.ca/index.php/cjlt/article/viewArticle/517/247

Cai, M. & Sims Bishop, M. (1990). Multicultural literature for children: Towards a clarification of the concept. In: A. H. Dyson & S. Gemishi (Eds.), *The need for story* (pp. 57-71). Urbana, Illinois: NCTE.

Cuban, L. (2001). Why are most teachers infrequent and restrained users of computers in their classrooms? In: J. Woodward & L. Cuban (Eds.), *Technology curriculum and professional development*. Thousand Oaks, CA: Corwin Press.

Darling-Hammond, L., Banks, J., Zumwalt, K., Gomez, L., Gamoran Sherin, M., Griesdorn, K., & Finn, L-E. (2005). Educational goals and purposes: Developing a curricular vision for teaching. In: L. Darling-Hamond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 169-200). San Francisco, CA: Jossey-Bass.

Désilets, A., & Paquet, S. (2005). Wiki as a tool for web-based collaborative story telling in primary school: A case study. *EdMedia 2005, World Conference on Educational Multimedia, Hypermedia & Telecommunications*. Montréal, Québec, Canada. June 27 - July 2 2005.

Downey, S., Wentling, R. M., Wentling, T., & Wadsworth, A. (2005). The Relationship between National Culture and the Usability of an E-learning System. *Human Resource Development International, 8*(1), 47-64.

Dwairy, M. A. (2006). *Counseling and psychotherapy with Arabs and Muslims: A culturally sensitive approach*. N.Y.: Teachers College Press.

Farabaugh, R. (2007). 'The Isle is Full of Noises': Using Wiki Software to Establish a Discourse Community in a Shakespeare Classroom. *Language Awareness*, *16*(1), 41-56.

Feiman-Nemser, S., & Floden, R. E. (1986). The cultures of teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 505-526). New York: Macmillan.

Gunawardena, C. N., Wilson, P. L., & Nolla, A. C. (2003). Culture and Online Education. In: *Handbook of distance education.* (pp. 753-776). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hall, A. (2006). Who's learning? Responding to the needs of a culturally diverse world of online learners. *Proceedings of the 23rd annual ASCILITE conference.* The University of Sydney. Retrieved April, 30, 2009, from: http://www.ascilite.org.au/conferences/sydney06/proceeding/pdf_papers/p138.pdf

Hall, E. T. (1966). *The hidden dimension*. Garden City, NY: Doubleday.

Hall, E. T. (1976). *Beyond culture*. Garden City, NY: Doubleday.

Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations, 10*, 301–320.

Kooy, M. (2006). *Telling stories in book clubs – Women teachers and professional development*. New York, NY: Springer.

Leuf, B. & Cunningham, W. (2001). *The Wiki Way: Collaboration and Sharing on the Internet.* Boston: Addison-Wesley.

Lyndsay, G. (2006, May). Using Wikis in schools. *Future lab innovation in education.* Retrieved April, 30, 2009, from: http://www.futurelab.org.uk/resources/documents/discussion_papers/Wikis_in_Schools.pdf

Matsumoto, D. (1996). *Culture and psychology.* Pacific Grove, CA: Brooks/Cole Publishing Company.

Meshar-Tal, H., & Tal-el-Hasid, E. (2006) Collaborative assignments in Wiki-based environment academic courses. *Al Ha Gova, 6*, 44-47 (Hebrew).

Morgan, M. (2004). Notes towards a Rhetoric of Wiki. *Paper presented at CCCC 2004*, San Antonio TX. Retrieved April, 30, 2009, from: http://mcmorgan.org/academics/papers_and_presentations/notes_towards_a_rhetoric_of/

Peleg, R. & Raslan, S. (2003). *An assessment of a teacher education program for minority students at the Oranim College of Education*. The College Research Authority, Oranim. (Hebrew).

Ravid, G. (2006). Wiki: New challenges in higher education. *Al Ha Gova, 6*, 40-43 (Hebrew).

Ravid, G. Kalman, Y. M., & Rafaeli, S. (2008).Wikibooks in higher education: Empowerment through online distributed collaboration. *Computers in Human Behavior, 24*, 1913-1928.

Robertson, I. (2008). Learners' attitudes to Wiki technology in problem based, blended learning for vocational teacher education. *Australian Journal of Educational Technology, 24*(4), 425-441.

Ruth, A. & Houghton, L. (2009). The wiki way of learning. *Australasian Journal of Educational Technology, 25*(2), 135-152.

Sumara, D. J. (2002). *Why reading literature in school still matters*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wang, L., & Beasley, W. (2008). The Wiki as a web 2.0 tool in education. *International Journal of Technology in teaching and learning, 4*(1), 78-85.

Watson, R. T., Boudreau, M-C., York, P. T., Greiner, M. & Wynn, D. E. (2008). Opening the classroom. *Journal of Information Systems Education, 19*(1). Retrieved April, 30, 2009 from: http://202.116.45.198/kcyjxl/jxtj/pdf/9/3.Opening%20the%20Classroom.pdf

Wheeler, S. & Wheeler, D. (2009). Using Wikis to promote quality learning in teacher education. *Learning, Media and Technology,* 34(1), 1-10.

Wild, M. (1999). Editorial: Accommodating issues of culture and diversity in the application of new technologies. *British Journal of Educational Technology*, *30*(3), 195-199.

Yildiz, S., & Bichelmeyer, B. A. (2003). Exploring electronic forum participation and interaction by EFL speakers in two web-based graduate-level courses. *Distance Education, 24*(2), 175–193.

Zaharna, R.S. (1995).Understanding cultural preferences of Arab communication patterns. *Public Relations Review, 21*(3), 241-255.

# The use of a digital dance mat for training kindergarten children in a magnitude comparison task

Ulrike Cress, Knowledge Media Research Center, Tuebingen, u.cress@iwm-kmrc.de
Ursula Fischer, Knowledge Media Research Center, Tuebingen
Korbinian Moeller, Psychological Institute, University of Tuebingen
Claudia Sauter, Knowledge Media Research Center, Tuebingen
Hans-Christoph Nuerk, Psychological Institute, University of Tuebingen

**Abstract:** Previous studies demonstrated that basic numerical skills reliably predict children's future mathematical performance. The spatial representation of numerical magnitude, represented in the form of a mental number line, seems to be of particular importance. Our training program for kindergarten children used a digital dance mat as input device that required children to move their whole body to respond in a magnitude comparison task. By employing such a spatial embodied training method, in a parallel randomized cross-over design, our study with 19 kindergarten children revealed a significant interaction between training condition and repeated exposure to items, implying that children improved more strongly in the dance mat than in the control condition. These results suggest that the use of digital media to train embodied spatial numerical skills may be more effective in basic numerical tasks such as magnitude comparison. We suggest that the involvement of embodied spatial codes, shared by the representation addressed by the task at hand, aids acquisition of task-relevant basic numerical skills.

## Basic numerical skills

Arithmetic competencies are important cultural skills, comparable to reading and writing being a fundamental requirement not only for school success, but also for coping with everyday life. At the beginning of their first year of schooling, children's arithmetic skills already differ markedly, and these differences do have long term consequences. It has been repeatedly demonstrated that preschool numerical skills are a good predictor of a child's later arithmetic performance (e.g., Duncan et al., 2007). Findings so far suggest that arithmetic skills and processing of numbers involve various basic numerical competencies that are based on specific types of representations. For instance, the *Triple Code Model* by Dehaene (Dehaene & Cohen, 1995; Dehaene, Piazza, Pinel, & Cohen, 2003) suggested three types of number representations: (1) the visual Arabic number form (written digits); (2) a verbal representation (written or spoken number words); and (3) an analogue or semantic representation of number magnitude (representing the quantitative meaning of numbers). Over the years, the Triple Code Model has been revised several times, adding among other changes a spatial representation of number magnitude in the form of a mental number line that is activated whenever a number is encountered (Dehaene et al., 2003; Nuerk, Graf, & Willmes, 2006). This is illustrated best by the so-called *SNARC* (*S*patial *N*umerical *A*ssociation of *R*esponse *C*odes) *effect* that is usually observed in parity judgement tasks as a systematic interaction between the side of response and number magnitude. In a large number of studies (e.g., Dehaene, Bossini, & Giraux, 1993; Nuerk, Iversen, & Willmes, 2004a; Nuerk, Wood, & Willmes, 2005; see Wood, Willmes, Nuerk & Fischer, 2008 for a meta analysis), subjects have been found to respond faster to small numbers by the left hand and faster to larger numbers by the right hand. Note that in parity judgement, the magnitude of the numbers as such is irrelevant to solving the task. Nevertheless, response patterns suggest that perception of a number automatically activates a spatial mental representation of its magnitude, which cannot be suppressed willingly. The development of the mental number line (as indicated by the presence of a SNARC effect) is regarded as a key requirement for later arithmetic achievement (e.g., Bachot, Gevers, Fias, & Roeyers, 2005) and a precise spatial representation of number magnitude was found to be associated with better actual mathematics achievement and a better ability to learn unknown arithmetical problems (Booth & Siegler, 2008). Therefore, training of such basic numerical representations may be a promising approach. However, before turning to the embodied spatial training program developed for the current study previous approaches on how to train arithmetic capabilities shall be reviewed briefly.

## Training arithmetic skills

Griffin, Case, and Siegler (1994) developed *Rightstart*, an intervention program for kindergartens. They were motivated by their observation that children from low-income families enter school with less understanding of mathematical concepts than their middle-class peers. In order to counteract this discrepancy and the resulting disadvantage in these children's later school career, the authors also developed versions of *Rightstart* (later renamed *Number Worlds)* for older age groups (Griffin, 2003). The program consists of game-like exercises on number lines, number comparison and the semantic representation of numbers. It also contains specifically developed board games that children can play in pairs or small groups (see also Ramani & Siegler, 2008). These board games train skills related to counting and visualizing the distance travelled by the figures, thus connecting the semantic and spatial representation of number magnitude. Evaluation studies confirmed a positive effect of the program on arithmetic capabilities of preschoolers (Griffin et al., 1994).

In 2003, Samara and Clements developed a computer program, *Building Blocks,* for children of kindergarten age (see also Clements, Battista, Samara, & Swaminathan, 1997; Clements & Battista, 1989; Clements & Meredith, 1993). The underlying instructional approach is genuinely constructivist and constructionist (cf. Papert, 1980): Children are expected to understand numerosity by manipulating objects systematically. The computer program allows a variety of manipulations and visualizations that support children's understanding of numbers. They can move objects – comparable to the board game – and this movement leaves traces that visualize distances and thus the spatial and semantic meaning of numbers (see also the training program *number race*; Wilson et al., 2006a; Wilson, Revkin, Cohen, & Dehaene, 2006b). Summative evaluation of this program confirmed that the children's numeric achievement improved by almost 1 standard deviation (Clements & Sarama, 2007).

## Embodied cognition

One hypothesis of the present study is that the development of the spatial magnitude representation can be supported by motor processes. This assumption is based on findings on *embodied cognition*. In traditional theories of cognitive science, perception and action planning were regarded as two separate processes, and cognitive processes were thought to be exclusively based on mental representations. The possibility of a connection between cognition and action was largely ignored for a long time and thus, influences of action-related processes on the perception of information, and the influence of perception processes on motor action were highly underestimated. However, more recent embodied cognition approaches specifically address these possible connections. For instance, Hommel, Müsseler, Aschersleben, and Prinz (2001) propose a framework that is based on the assumption that the contents of cognition and action plans are encoded in a common architecture of representation, and integrated in a joint task-oriented network.

The most basic elements of this *Theory of Event Coding* (TEC) are so-called *feature codes* that represent all features of an event. These refer to different aspects or phenomena of a specific task which they represent. Certain *features* do not refer specifically and exclusively to one stimulus or response, but process sensory input from various different sensory systems and modulate activities of different motor systems. According to TEC, an event will activate several different feature codes, each of which modulates one specific aspect of perception *and* action. Processes of perception and action interact in the jointly shared representational medium of encoding when *features* of perception and motor functions overlap. Feature codes of both perceived information and action plans are then integrated in the same representational medium when they become part of an *event code*. Hommel and colleagues assume that event codes are accumulations of feature codes and that their temporary integration will depend on the respective context and task. Many features of events in the environment exist in different sensory modalities, so restrictions of one modality may be compensated by taking into account and integrating information from another modality. A task is thus solved more easily the more features stimulus and response have in common. On this basis, we can make assumptions about the supportive role of motor skills in learning. We expect target-oriented motor movements, combined with external representations of stimuli (in this case: of numerical magnitude) to lead to an integration of the respective *feature codes*. Thus, a joint representational architecture for number perception and motion will be built. While Hommel and colleagues inferred evidence for their theory from various empirical studies they did not take into consideration the specific connection between number magnitude processing and motor skills, or *embodied numerosity*. However, various empirical findings do suggest that such a connection exists. In many cultures, parts of the body – mostly hands and fingers – are used to represent numbers (cf. Menninger & Broneer, 1992). Furthermore, even adult number processing appears to be closely linked to finger counting (e.g., Domahs, Krinzinger, & Willmes, 2008; Noël, 2005; Wood, Willmes, Nuerk, & Fischer, 2008; Fischer, 2008) and the motor system of grasping (e.g., Badets, Andres, Di Luca, & Pesenti, 2007; Moretto & DiPellegrino, 2008).This indicates that numerosity is not only represented mentally, but also bodily, and that a corresponding connection between motor skills and number processing exists.

## Media for embodied cognition

Different new digital media provide input formats that enable an *embodied interaction* (Romero, Good, Robertson, du Boulay, Reid, & Howland, 2007; Dourish, 2001). Such interfaces have been developed in recent years mainly for computer games. For example, *digital dance mats* require subjects to move their whole body by stepping on different fields of the mat in a specific succession. Games using dance mats have become widespread as part of the Sony Playstation® 2 (PS2) or Microsoft XBOX. Low-price versions of dance mats are now available in supermarkets, and can be connected to any computer via USB. In addition, the so-called *Wii* hit the market in 2006. This gaming console includes a handheld pointing device that does not work by activating defined fields, but by moving a controller which contains sensors detecting movements, thus registering the player's position and actions. More recent systems of the same type detect movements and gestures via web cams (Hoysniemi, 2006; Moeslund & Hilton, 2006; Fitzgerald et al., 2006).

Comparable interfaces for embodied interaction have already been applied for games involving motion (e.g., Reidsma, Nijholt, Poppe, Rienks, & Hondorp, 2006), but also for training of physical fitness (e.g., Hartnett, Lin, Ortiz, & Tabas, 2006) and health education programs (e.g., Watters et al., 2006). In the U.S. dance mats were used in more than 760 schools in the context of physical education, hoping to achieve a long-term effect on the students' health behavior (Business Wire, 2006). To our knowledge, dance mats have never been used in a specifically instructional context before. We believe, though, that this input device that is including spatial and motor input components has potential for training the spatial representation of numbers. Additionally, digital dance mats have a high motivational potential. Children have fun training and moving with them, and they are thus particularly suitable for use in kindergarten and primary school.

## The current study

Based on the considerations described above, we expected that an embodied spatial-numerical training realizing shared features in both presentation and response format should be more effective in leading to a more pronounced decrease in error rates and response latencies. The specificities of the training program we developed will be described below.

We trained children on a magnitude comparison task in two conditions using different presentation and response formats. In the experimental (dance mat) condition, additional presentation of a number line and a spatial response format that required motor input supported children's performance. We designed the control (tablet PC) condition so that neither presentation nor response format involved explicit spatial information. In both conditions, we used two types of presentation material: Arabic numerals and assemblies of squares, which were presented inside a larger square to be perceived as an entity. Prior to the task, children were instructed not to count the squares in the assemblies but to estimate how many there were. To avoid giving the children perceptual cues, non-numeric factors of the assemblies were kept constant (as proposed by Xu & Spelke, 2000). For example, the white surface covered by black squares and thus, the luminosity of the stimuli was held constant in all items of the assembly variation.

### Experimental condition: Dance mat

At the beginning of each trial, children stood on the central field of the dance mat. Tasks were projected onto the floor with a data projector directly in front of them (see Figure 1). The presented tasks consisted of lines that had either numbers (*0* and *10* or *0* and *20*) or square assemblies (0 and 10 or 0 and 20 squares) marking the endpoints. Between the endpoints, one number/square assembly was marked as a standard for comparison. The children had to compare the magnitude of a second number/assembly that was presented above the line with the magnitude of the standard and to judge whether its position on the line was left or right of the standard. Responses were given by stepping on the left field of the mat with both feet when the number's/assembly's location was to the left and on the right field when it was to the right of the standard.

Figure 1: Experimental set-up in the dance mat condition.

## Control condition: Tablet PC

In the control condition, the same magnitude comparison task was administered in a different setup. The control condition's function was to provide a comparable task that involved new media as well, but did not provide as much of a strong connection between action and perception. Children were shown two numbers that were vertically aligned on the monitor of a tablet PC. In contrast to the dance mat condition, children merely had to touch the bigger number with an electronic pen. The same stimuli in pseudo-randomized order were used in the control and experimental condition. Figure 2 shows the stimulus formats used in both conditions (separated by type of presentation material) on the example of the comparison between 3 and 7.

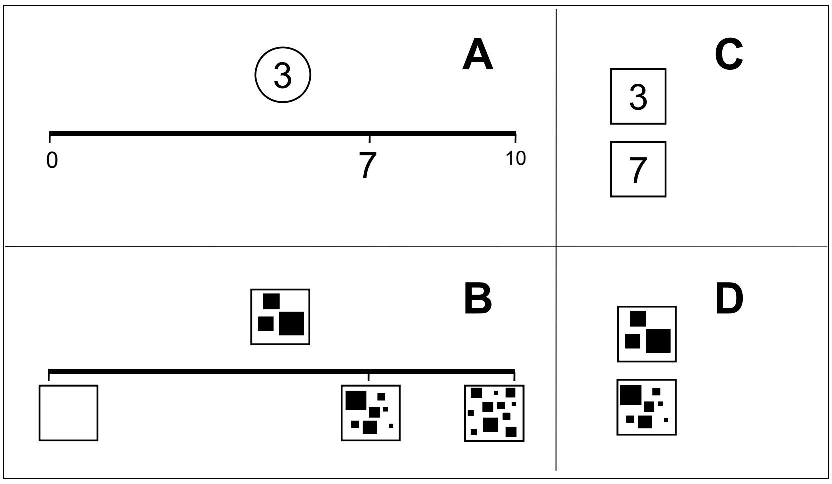



Figure 2: Examples of stimuli in both types of presentation material in the dance mat condition (with **A.** Arabic numbers and **B.** square assemblies) and in the tablet PC condition (with **C.** Arabic numbers and **D.** square assemblies).

## Procedure and Design

We chose a parallel randomized cross-over design to provide all children with the possibility of training with the dance mat, meaning that half of the children first received dance mat and then tablet PC training, while for the other half the order of training conditions was reversed.

The data presented in this paper were part of a large scale study. For reasons of brevity, the current article will address the results of the magnitude comparison task only. In this large-scale study, we also varied whether numbers/square assemblies had to be compared to a variable or fixed standard (with fixed standard magnitudes being either 5 or 10). In the scope of this paper, we will focus on items on which magnitudes had to be compared to a variable standard, because these items – unlike items with a fixed standard – were presented twice each and allowed for an evaluation of children's improvement from the first to the second time of an item's presentation. This way, we could analyze whether children improved similarly over training in either condition or whether one condition yielded a higher improvement than the other.

Each child received three training sessions with the dance mat and three sessions with the tablet PC, which consisted of 64 to 72 items each. The items of the first training session of either condition consisted of magnitudes ranging from 0 to 10, the items of the second session from 0 to 20, and every third session comprised half of the items of the first and half of the items of the second session (i.e., comparisons to a variable standard only). The six training sessions were carried out individually with each child on six different days within a period of three weeks. Training sessions took approximately 10 to 15 minutes per child and were administered before noon in rooms of the kindergarten. When a child missed or could not participate in a training session, the missed session was – if possible – carried out on some other day.

## Participants

Preschool children from two kindergartens participated in the study. The parents of 27 children gave their written consent. Because of prolonged illness or absence, 5 children dropped out of the study. Children were randomly assigned to the two conditions, 11 children were first trained on the dance mat and 11 started with the tablet PC. Subsequently, 3 children's data sets had to be excluded from the analysis due to absence in one training session that could not be carried out at a later point or poor understanding of task instructions, leaving the complete data sets of 19 children – 8 girls, 11 boys –, aged between 60 and 79 months, for analyses. Mean age of the children included in the analysis was $M = 70.74$ ($SD = 6.06$) months and did not differ between the two experimental conditions.

## Results

To analyze the learning effects of training with the dance mat as compared to training with the tablet PC, we compared children's improvement in accuracy (rates of correctly solved items) and reaction times between conditions on items that were presented twice. Improvement was computed by subtracting performance of all children at the first time of presentation from their perfomance at the second time of presentation separately for each item. Since means of reaction times differed between the two conditions due to the very different response formats, z-standardized reaction times were used to ensure comparability of conditions. Reaction times were only computed for correct trials. Children's rates of correctly solved items were arcsine transformed prior to all analyses to approximate normal distribution. We then conducted Analyses of Variance comparing each child's improvement in accuracy and reaction time between the two conditions ($F_1$). Furthermore, we analyzed the data across items ($F_2$). Where it was necessary, one-sided *t*-tests (i.e., $t_1$ and $t_2$, respectively) were used to evaluate our directed hypotheses.

### Children's Performance

First, we conducted ANOVAs on children's improvement in accuracy and on their decrease in reaction time with *condition* and *material* (digits or square assemblies) as within subject factors. We further tested for the predicted differences between the two conditions by means of one-sided *t*-tests. As expected, we found that children's z-standardized reaction times decreased significantly more in the dance mat than in the tablet PC condition [$t_1(18) = 6.19$, $p < .001$, tested one-sided; ANOVA: $F_1(18) = 24.14$]. Likewise, there was a significant main effect of condition on rates of correctly solved trials [$t_1(18) = 2.14$, $p < .05$; $F_1(18) = 2.94$], indicating that children's increase of accuracy from the first to the second time of presentation was larger in the dance mat than in the tablet PC condition (see Figure 3). We found a marginally significant main effect of material on accuracy [$F_1(18) = 3.18$, $p = .08$] that would have reached significance if tested one-sided, but no significant effect of material on reaction times [$F_1(18) = 2.32$, $p = .14$]; and there was no significant interaction between material and condition in either accuracy or reaction times [both $F_1(18) < 1$]. This lack of significant interactions implies that the differences between the conditions were not mainly caused by one type of material, but that children improved similarly on items consisting of both digits and square assemblies.

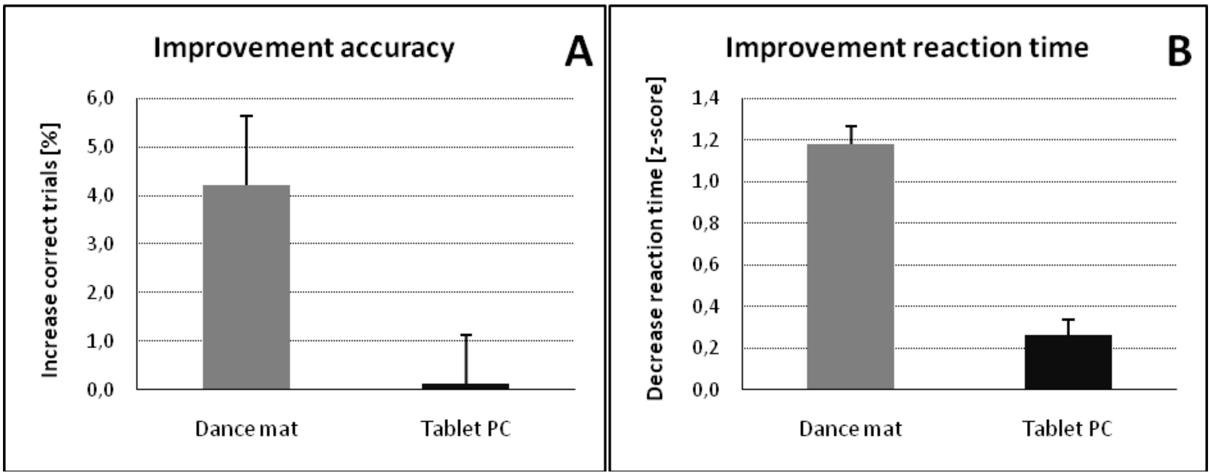



Figure 3: Comparison of children's improvement in **A.** accuracy and **B.** reaction time in the dance mat and tablet PC condition.

### Item analysis

For the Item analysis, repeated measures ANOVAs were conducted, this time with *condition* as the repeated measures factor and *material* as the between subject factor. Again, hypothesized differences were examined by use of one-sided *t*-tests. We found a significant main effect of condition on rates of correctly solved trials [$t_2(67) = 2.66$, $p < .01$; ANOVA: $F_2(1,67) = 5.54$], as well as on reaction times [$t_2(67) = 4.20$, $p < .001$; $F_2(1,67) = 16.03$]. *Material* was found to have a marginally significant main effect on accuracy [$F_2(1,67) = 3.48$, $p = .06$] and no significant main effect on reaction times [$F_2(1,67) = 1.74$, $p = .19$]. Again, material and condition did not interact significantly in either accuracy or reaction times [both $F_2(1,67) < 1$].

### Subjective evaluation

At the end of the study, we interviewed parents and kindergarten staff on their impression of the training. Both groups agreed that the children enjoyed both the dance mat and tablet PC training, but that children had more fun on the dance mat. Most children talked about the project at home and in the kindergarten; some of them spent more time on mathematical tasks than before. Parents and kindergarten teachers had the impression that the level of difficulty was adequate for the children's age and abilities, and the kindergarten staff approved of the idea of using the dance mat for future projects.

## Discussion

The present study showed that the magnitude comparison task can be trained in kindergarten children and that improvement in the magnitude comparison task is larger when this task is trained using an embodied spatial-numerical approach. Performance of children improved significantly more in the dance mat condition than in the tablet PC condition on both reaction times and accuracy. This superiority of the dance mat training was found in analysis of children's performance and also persisted when an item analysis was conducted. These results provide further support for Hommel et al.'s (2001) theory of event coding, indicating that the mental number line representation is indeed more strongly activated when presentation and response format share more features. In the present case mainly spatial attributes are enforced by motor movements and the presentation of the to-be-compared numbers on a number line. Thus, in accordance with Hommel et al., features of stimulus and response seem to be processed as parts of the same event code. Also, the lack of a significant interaction between condition and material (digits or square assemblies) speaks for a generalizability of the positive effects of the dance mat training across different types of stimulus material.

It should also be noted that our computations yielded marginally significant main effects for material (digits or square assemblies) on accuracy that would have reached significance had we employed a one-sided test. However, we did not have any expectations as to which material would produce higher training effects than the other. A trend was found for means of improvement to be higher on items consisting of digits than on items consisting of square assemblies. This might be due to the distinct nature of Arabic digits that allowed for children to remember them more accurately than square assemblies, which children were instructed to process only in an approximate manner. Thus, children might have remembered Arabic digits, but not patterns of square assemblies.

One might argue that the tasks we used were too demanding for kindergarten children, since the German curriculum for the last year of kindergarten states that children should be able to recognize the regular dice patterns. At the time they enter school, the children are thus expected to be familiar with the numbers from 1 to 6. The tasks used in the training program of the current study presenting magnitudes from 0-20 exceeded this level by far. However, since chance level of success was 50 percent on all the tasks, children did not experience failure too often keeping their motivation at a very high level. Furthermore, the fact that significant differences in learning outcome were observed between the two conditions when items were presented only twice argues for the high potential of the dance mat as a training medium for basic numerical skills.

A differentiation of which processes precisely differentiate the two training conditions should be the aim of further studies. In the present experiment, the dance mat and tablet PC condition differed in various features of both perception and action. The dance mat condition was combined with presenting the tasks in number line form, and children had to respond by using motor skills to move their whole body to one side or the other. In the tablet PC condition, items were presented without external representation of the number line, and the response format did not ask for any movement to the left or right. So, from the present study, the question remains whether it were either differences of perception *or* action that led to the performance difference between dance mat and tablet PC condition, or if it was the combination of these two features.

Furthermore, it would be of interest whether the improvements in children's performance on the magnitude comparison tasks may be transferable to other domains of numerical knowledge. We examined children's performance in pre- and posttest transfer measures, namely a number line estimation task and a standardized arithmetic test battery. These results, however, will be published separately and will therefore receive no further discussion in this paper. Seeing as children improved significantly more in both transfer measures in the dance mat than in the tablet PC condition, these transfer effects do enforce the assumption that children profit more from the dance mat than training.

Taken together, the present study showed that a digital dance mat can be used effectively for training of the spatial representation of number magnitude. Bearing in mind that the spatial representation of number magnitude is an important basic competence that reliably predicts later mathematics achievement (Booth & Siegler, 2008), training of this competency would be desirable in preschool education. Moreover, since subjective ratings clearly favoured the dance mat over the tablet PC as well, we conclude that exercise with the dance mat presents a possibility of such training that children experience as a highly motivating game.

On a general level, this study indicates that the theory-guided use of modern media improves learning even in basic tasks and thus supports the idea that multi-media tools can be an important mediator of learning success even in kindergarten children. Moreover, this study lends support to the idea that the inclusion of bodily experiences to represent abstract concepts may not only be helpful to represent those abstract concepts but that embodied cognition also aids their acquisition.

## References


Bachot, J., Gevers, W., Fias, W., & Roeyers, H. (2005). Number sense in children with visuospatial disabilities: orientation of the mental number line. *Psychology Science 47*, 172-183.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79,* 1016-1031.

Business Wire (2006). Abgerufen am 11. August 2009. Konami Digital Entertainment and the State of West Virginia develop groundbreaking program that brings the health benefits of Dance Dance Revolution to every school in the State: http://findarticles.com/p/articles/mi_m0EIN/is_2006_Jan_25/ai_n26735500.

Clements, D. H., & Battista, M.T. (1989). Learning of geometric concepts in a logo environment. *Journal for Research in Mathematics Education, 20,* 450-467.

Clements, D. H., Battista, M. T., Samara, J., & Swaminathan, S. (1997). Students' development of length concepts in a logo-based unit on geometric paths. *Journal for Research in Mathematics Education, 28* (2), 70-95.

Clements, D. H., & Meredith, J.S. (1993). Research on logo: Effects and efficacy. *Journal of Computing in Childhood Education, 4,* 263-290.

Clements, D. H., & Samara, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38*, 136-163.

Dehaene, S., & Cohen, L. (1995). Toward an anatomical and functional model of number processing. *Mathematical Cognition, 1*, 83-120.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*, 371-396.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20,* 487-506.

Dourish, P. (2001). *Where the Action is: The Foundations of Embodied Interaction.* Cambridge, MA: MIT Press.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L., Feinstein, L., Engel, M., & Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex, 44,* 386-392.

Fitzgerald, D., Foody, J., Kumar, D., Kelly, D., Ward, T., Markham C., & Caulfield B. (2006). Integration of kinematic Analysis into Computer Games for Exercise. In *CGAMES 2006 - 9th International Conference on Computer Games: AI, Animation, Mobile, Educational and Serious Games* (pp. 24-28).

Griffin, S. (2003). Number Worlds: A research-based mathematics program for young children. In D. H. Clements, & A. DiBiase (Eds.), *Engaging Young Children in Mathematics: Findings of the 2000 National Conference on Standards for Preschool and Kindergarten Mathematics Education* (pp. 325-342). Hillsdale, NJ: Erlbaum Associates, Inc.

Griffin, S., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice* (pp. 1-50). Cambridge, MA: MIT Press/Bradford Books.

Hartnett, J., Lin, P., Ortiz, L., & Tabas, L. (2006). A responsive and persuasive audio device to stimulate exercise and fitness in children. *CHI '06 Extended Abstracts on Human Factors in Computing Systems.* New York: ACM.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849-937.

Hoysniemi, J. (2006). International survey on the dance revolution game. *ACM Computers in Entertainment, 4* (2).

Menninger, K., & Broneer, P. (1992). *Number Words and Number Symbols: A Cultural History of Numbers.* London: MIT Press.

Moeslund, T. B., & Hilton, A. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 90-126.

Moretto, G., & di Pellegrino, G. (2008). Grasping numbers. *Experimental Brain Research, 188,* 505–515.

Noël, M. P. (2005). Finger gnosia: A predictor of numerical abilities in children? *Child Neuropsychology, 11*, 413-430.

Nuerk, H.-C., Graf, M., & Willmes, K. (2006). Grundlagen der Zahlenverarbeitung und des Rechnens [Foundations of number processing and calculation]. *Sprache, Stimme, Gehör: Zeitschrift für Kommunikationsstörungen, Schwerpunktthema Dyskalkulie, 30*, 147-153.

Nuerk, H.-C., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness association of response codes) effect. *Quarterly Journal of Experimental Psychology, 57*, 835-863.

Nuerk, H.-C., Wood, G., & Willmes, K. (2005). The universal SNARC effect: The association between number magnitude and space is amodal. *Experimental Psychology, 52,* 187-194.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful ideas*. New York: Basic Books.
Ramani, B. G., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board Games. *Child Development, 79 (2)*, 375-394.
Reidsma, D., Nijholt, A., Poppe, R., Rienks, R., & Hondorp, H. (2006). Virtual rap dancer: Invitation to dance. *In CHI '06 Extended Abstracts on Human Factors in Computing Systems* (pp. 63–266). New York: ACM.
Romero, P., Good, J., Robertson, J., du Boulay, B., Reid, H., & Howland, K. (2007). Embodied interaction in authoring environments. In D. Ramduny-Ellis, J. Hare, S. Gill, & A. Dix (Eds.), *Proceedings of the Second Workshop on Physicality* (pp. 43-46). Lancaster, UK: UWIC Press.
Samara, J., & Clements, D. (2003). Building blocks of early childhood mathematics. *Teaching Children Mathematics, 9* (8), 480-484.
Watters, C., Oore, S., Shepherd, M., Abouzied, A., Cox, A., Kellar, M., Kharrazi, H., Liu, F., & Otley, A. (2006). Extending the Use of Games in Health Care. In *Proceedings of the Hawaii International Conference on System Science (HICSS '06),* Kauai, HI. (CD-ROM Proceedings)
Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S., Cohen, L., & Cohen, D. (2006a). Principles underlying the design of The Number Race, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2:19*.
Wilson, A. J., Revkin, S., Cohen, D., Cohen, L., & Dehaene, S. (2006b). An open trial assessment of The Number Race, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2:20*.
Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychology Science Quarterly, 50*, 489-525.
Xu, F. & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition 74,* B1-B11.

## Acknowledgments

We would like to thank Martina Bientzle for her help in this study, especially concerning data acquisition. Furthermore, we wish to thank the kindergartens Stuttgart-Birkach and St. Martin in Tuebingen-Hirschau for their helpful and constructive cooperation in this study.

# Teacher-education students' views about knowledge building theory and practice

Huang-Yao Hong, National Chengchi University, Taiwan, hyhong@nccu.edu.tw
Fei-Ching Chen, National Central University, Taiwan, fcc@cc.ncu.edu.tw
Ching Sing Chai, NIE, Nanyang Technological University, Singapore, chingsing.chai@nie.edu.sg
Wen-Ching Chan, National Chengchi University, Taiwan, 96152020@nccu.edu.tw

**Abstract**. This study investigated the effects of engaging students to collectively learn and work with knowledge in a computer-supported collaborative learning environment called Knowledge Forum on their views about knowledge building theory and practice. Participants were 24 teacher-education students who took a required course about theory and practice in teaching. Data mainly came from (1) student discourse recorded in a Knowledge Forum database, (2) a survey that examined students' views about knowledge building, and (3) interviews with regard to students' perceived barriers to implementing knowledge building theory in teaching. Findings suggest that with sustained discourse to construct their collective understanding of the relationships between theory and practice in teaching for a semester, the participants were able to attain more informed and practical views about knowledge building theory. In addition, students' perceived barriers to implementing knowledge building in teaching were identified and strategies to help overcome these barriers discussed.

## Introduction

Teaching has been viewed as a craft (Bereiter, 2002). As commonly observed in the classroom, most teachers tend to pursue "best practices" by practicing their teaching according to some known theories, and they are less inclined to go beyond "best practices" and assume the role of theory-building for their practice (Hargreaves, 1999; Sawyer, 2004). Recent literature, however, emphasizes the importance of viewing teaching as a knowledge-building enterprise (Hargreaves, 1999; Zhang, Hong, Teo, Scardamalia & Morley, 2008; Scardamalia, 2002). Related concepts have been introduced to support this idea, for example, creative teaching (Sawyer, 2004), adventurous teaching (Cohen, 1989), and teaching as progressive problem solving (Bereiter & Scardamalia, 1993) or as a sustained design process (Hong, Zhang, Teo, & Scardamalia, 2009). Yet, the idea of education as a progressive science and teaching as knowledge building is still new to most teachers (Sawyer, 2006).

One way to help teachers to develop a deeper conceptual understanding of teaching as a process of knowledge-building may be to engage them in the actual "knowledge-building" practice (Hargreaves, 1999; Hong & Sullivan, 2009). Knowledge-building is a social process focused on the production and continual improvement of ideas of value to a community (Bereiter & Scardamalia, 2003). The epistemological position underlying the knowledge building pedagogy is Popper's (1972) construct of World 3. Other than World 1 (the physical world) and World 2 (the subjective world inside the mind), Popper postulates a World 3 that is constituted of conceptual artifacts. The ideas and theories created by knowledge workers such as scientists, engineers and architects are among the conceptual artifacts. These theories and ideas, once created, have a life of their own in that they can be improved and transformed by people who interact with them. They are treated as tentative theories that should be subjected to error elimination under Popper's schema for the search for truth. In other words, all created knowledge is open to further inquiry and improvement. This epistemological stance is translated directly into the practice of treating all knowledge as ideas and as improvable in a knowledge-building community (Scardamalia, 2002). Bereiter (1994) argues that school focused on changing students' mind (ie, World 2) and neglected the enculturation of students' competencies to work in World 3. Arguably, teachers are unaccustomed to the ways of building knowledge as professionals, much less developing such competencies among students (Hong, Scardamalia, & Zhang, 2007).

To facilitate the process of knowledge building, a set of 12 knowledge-building principles have been conceptualized (Scardamalia, 2002). These principles have evolved over the last two decades: from an earlier focus on transformative discourse (Bereiter & Scardamalia, 1987), intentional learning (Scardamalia & Bereiter, 1991), and creative expertise as progressive problem solving (Bereiter & Scardamalia, 1993), to the most recent 12 knowledge building principles (Scardamalia, 2002). These 12 principles represent some innovative, pedagogical know-how to help transform a traditional class into a knowledge building community. They include (1) *Real Ideas, Authentic Problems*; (2) *Idea Diversity;* (3) *Improvable Ideas*; (4) *Epistemic Agency*; (5) *Community Knowledge, Collective Responsibility;* (6) *Democratizing Knowledge*; (7) *Symmetric Knowledge Advance*; (8) *Pervasive Knowledge Building*; (9) *Constructive Uses of Authoritative Sources*; (10) *Knowledge Building Discourse*; (11) *Concurrent, Embedded, Transformative Assessment*; and (12) *Rise Above* (see

Scardamalia, 2002, for more details). Fundamentally, knowledge building principles are designed to reconceptualize the behaviors of and relationships between three essential knowledge-building entities: the idea, the agent, and the community. For example, the principle of *Real Ideas, Authentic Problems* highlights the importance of viewing student ideas as conceptual artifacts (Bereiter, 2002) that are as real as things touched and felt, and that knowledge problems arise from efforts to understand the world and the ideas of other collaborators in the community, leading to problems of understanding that are quite different from textbook problems and puzzles. The principle of *Epistemic Agency* underscores that participants deal with the full range of knowledge problems (goals, motivation, evaluation, long-range planning, etc.), including knowledge problems normally left to teachers or managers. And the principle of *Community Knowledge, Collective Responsibility* emphasizes that contributions to shared, top-level goals of the community are prized and rewarded as much as individual achievements; team members produce ideas of value to others and share responsibility for the overall advancement of knowledge in the community. These principles represent design ideals and challenges that set the stage for the community's work in sustained knowledge advancement (Bereiter & Scardamalia, 2003), which is very different from conventional classroom work defined by pre-specified procedures, clear scripts and rules, or any highly-structured, ritualistic learning activities that represent fixed rather than improvable classroom procedures (cf. Hong & Sullivan, 2009).

A growing body of evidence has suggested that it is important to consider teachers' epistemological views since such views will influence classroom performance (Chai, Hong, & Teo, 2009; Pajares, 1992; Richardson, Anders, Tidwell & Lloyd, 1991; Wilson, 1990). The aforementioned principles represent essential concepts underlying knowledge building as a theory of knowing and as a way to transform traditional teaching practice. In order to help prospective teachers develop a more informed view of knowledge building theory and practice, instead of employing traditional direct teaching, the present study engaged the participants in self-initiated and self-directed knowledge work in a knowledge building environment (Hargreaves, 1999; Hong, Zhang, Teo, & Scardamalia, 2009). In particular, we are interested to find out whether knowledge-building environment and technology affects students' learning processes and outcomes. In terms of processes, we looked into participants' online performance patterns, social interaction patterns, and patterns in relation to their reflective understanding of the relationships between theories and practices in teaching. In terms of outcomes, we looked into pre-post changes in students' views about the importance and feasibility of knowledge building, and their perceived barriers to implementing knowledge building in class.

## Method

### Participants and context

The present research was conducted in a university course titled "Integrating Theory and Practice in Teaching" in Taiwan. The course was offered by the university's Center of Teacher Education to teacher-education students as its last required course before they start their teaching practicum. The university is ranked as one of the best universities in the nation. As such, the students enrolled in the subject university are all academically high-achievers. Based on the test results of the national Basic Competence Test for Senior High School Students (BCTSHSS), in order to enroll in this university, a student's test scores in BCTSHSS need to be ranked above the 95th percentile nationwide. Participants in the present study were 24 teacher-education students (14 females) who were planning to teach at the high-school level in the near future. Their ages range from 21 to 29 (M=24; SD=2.3).

### Instructional design and online knowledge building environment

By engaging students in building knowledge in Knowledge Forum, the two main insructional goals were: (1) to help students better understand the complex relationships between theories and practices in teaching; and (2) to help students develop a more informed and practical view about knowledge building. To these ends, an invited talk about knowledge building theory, pedagogy, and principles, and a tutorial workshop about how to use Knowledge Forum for knoweldge building were given in the beginning of the semester. The basic design features and functions of Knowledge Forum were demonstrated to students, for example, how to create a note or a "view" (i.e., virtual spaces for collaborative discourse among community members) and how to "build-on" to an existing note. Other major instructional activities included: (1) a weekly reading assignment in which students (a) reviewed literature related to various teaching theories, and (b) read teachers' interview transcripts in which in-service teachers share their successes and challenges encountered in their daily teaching practice; (2) an invited guest speaker (i.e., a veteran teacher) shared his personal teaching experiences; and, (3) most importantly, sustained online peer discussion about the relationships between theories and practices in teaching.

The technology platform used to support peer discourse is Knowledge Forum--a multimedia community knowledge space (Scardamalia, 2003), in which participants spend extensive time collectively constructing their knowledge. They contribute their ideas in the form of notes. The Knowledge Forum environment also enables participants to co-author notes, build-on, reference (i.e., citation excepted from other community members' notes), and annotate the work of others, set problem fields and add keywords, and create "rise-above" notes that bring greater coherence to the contents of the knowledge space. All these features are designed as different means to foster collaboration in depth. For example, the "rise-aboves" allow users to gather theories and ideas that have already been presented, synthesize these old ideas and point out new challenges to understandings. Operations such as reading, referencing, editing, rise-above etc. are recorded automatically in the database, and can be summarized statistically by means of an Analytic Toolkit (Burtis, 2002). The Knowledge Forum technology designs—in line with the overarching commitment to continual knowledge improvement—allow students to exchange ideas and continuously improve them.

## Study design and data sources

This research employed a mixed-method design. The rationale for using such a design is that "the quantitative data and results provide a general picture of the research problem; more analysis, specifically through qualitative data collection, is needed to refine, extend, or explain the general picture" (Creswell, 2005, p. 515). Data sources mainly came from: (1) students' online discourse which was recorded as notes in a Knowledge Forum database, (2) a survey, and (3) interviews. We describe each in detail below. First, a descriptive analysis and a social network analysis (SNA) were performed on the recorded dataset in the Knowledge Forum to describe in general the overall online knowledge building process. In addition, all notes in the Knowledge Forum database were content-analyzed to examine if students gained a deeper understanding of the relationships between theories and practices in teaching. To do so, an open-coding procedure (Strauss & Corbin, 1990), using the note as the unit of analysis, was adopted. A two-level coding scheme based on Anderson & Krathwohl's (2001) revision of Bloom's (1956) Taxonomy of Educational Objectives was adopted. The two levels are: (1) lower-level cognitive activity/responsibility (including remember, understand, apply teaching theories), and (2) higher-level cognitive activity/responsibility (including analyze, evaluate, and create teaching theories). Two researchers repeatedly read and re-read all students' notes and then categorized each note into a level. An inter-coder agreement was computed to be 0.89 (with all differences resolved by discussion). Table 1 shows the coding scheme. To determine whether there were any changes in terms of students' discoure levels, the whole semester was divided into two stages: an early and a later knowelge-building stage (using the midterm exam as a point of seperation). A Chi-square was computed to decide if there were any differences between the two stages in terms of the discoure levels.

Second, a survey that measures students' views about the importance and feasibility of knowledge building was administered in the beginning and at the end of the semester as a pre-post assessment. This survey was designed by the authors to assess participants' mindset about the importance and feasibility of knowledge-building. There are 12 items in this survey, each is represented by a knowledge-building principle (see Scardamalia, 2002, for details). Using subjects (N=22) from another teacher education program of a comparable university, the Cronbach Alpha reliability estimates were calculated to be .87 (for the "importance" dimension) and .74 (for the "applicability" dimension). All items in both surveys employed a 5-point Likert scale (1=strongly disagree; 5=strongly agree). T-tests were conducted to see if there were pre-post changes in students' views.

Third, an approximately one-hour long interview was conducted as a follow-up investigation to further explore the perceived barriers and challenges among the teacher-education students who have expressed concerns about implementing knoweldge building in their future teaching. Six (out of total 24) participants who rated knowelge building as important but less feasible in their surveys were approached and they agreed to participate in the follow-up interviews. The interview data were transcribed verbatim and qualitatively used to help uncover some major barriers to implementing knowledge building.

Table 1. Coding scheme for analyzing relationships between theory and practice in teaching

| Level | Focus | Description | Example (Translated from Chinese) |
|---|---|---|---|
| Lower-level cognitive activity/ responsibility | Remember/ understand | Teachers should know and understand theories; | Teacher should understand some basic theories, such as behavioral learning and collaborative Learning. |
| | Apply | Teachers should be able to apply theories in teaching | I think teachers need to apply different theories in different conditions. |
| Higher-level cognitive activity/ responsibility | Analyze/ Evaluate | Teachers should be able to analyze theories and practice | Experience and theory are like Na-Kon-Hsin-Fa [a type of Chinese Kung Fu]. After one masters some theories, they can help supplement and/or be integrated into one's own personal teaching experience. |

| | | |
|---|---|---|
| Create | Teachers should be able to improve and even create theories | Teacher's experiences and self-reflection can help with the development of new theories. |

## Results and Discussion

### Knowledge building practices

#### Online contribution patterns

The overall online activity and performance in this community is shown in Table 2. Throughout the whole semester, the participants contributed a total number of 625 notes with a mean number of 26.04 (SD=6.69) notes being generated per person. In addition, Table 2 also shows other related online knowledge-building measures recorded in this community, including number of note revisions, number of keywords in notes, and number of build-on notes generated, and number of rise-above notes created. Overall, the online activities were substantive as compared with our previous study (see, Hong & Lin, accepted; Chai & Khine, 2006). Nevertheless, while these behavioral measures gave a general picture of how participants worked online in this database, they do not tell much about how participants actually interacted with one another. To better understand the social dynamics in the community, a social network analysis (SNA) focusing on network density was conducted.

Table 2. Descriptive analysis on individual online knowledge-building activities

| Online activity | Mean | SD |
|---|---|---|
| No. of notes created | 26.04 | 6.69 |
| No. of note revisions | 8.5 | 7.0 |
| No. of keywords in notes | 6.6 | 4.21 |
| No. of build-on notes created | 10.2 | 4.45 |
| No. of Rise-above notes created | 1.1 | 0.81 |

#### Online interaction patterns

SNA was conducted to investigate interaction patterns in the community by using the automatic assessment tools embedded in the Knowledge Forum. Table 3 shows the overall interactive and collaborative patterns in the community throughout the whole semester, using two indicators that are available in the Knowledge Forum: passive "note-reading" and active "note-linking" (including build-on notes, references, and annotations). It also shows detailed results of participants' interactions in two knowledge-building stages (using the mid-term exam as a point of seperation). In this particular analysis, density is defined as the proportion of connections in a network relative to the total number possible. The higher the number of the density is, the stronger the social dynamics of a community is implied. An intention of adopting the knowledge-building practice in this course was to transform the traditional knowledge-transmission mode of learning into a knowledge-construction mode to engage these students in collective problem-solving and knowledge work. Therefore, it was expected that the students should collaborate more as they progress. As can be seen, there was an increasing trend of social interactions as reflected by the measures of density recorded online for this community from the early to the later knowledge building stages, especially in terms of note-linking (which inlcude build-on, reference, and annotation). Lipponen Rahikainen, Lallimo, and Hakkarainen (2003) regarded a social network density of .39 for students building-on each other online messages as adequate. In this study, the social network density for building-on at the end of course is 44.2%. The findings indicate a satisfactory level of social interaction in this community.

To further understand the quality of learning in this community, we content-analyzed students' notes. In so doing, we illustrate the processes of how they actually learn and deepen their understanding towards the pre-determined teaching goal, which was to better understand the relationships between theories and practices in teaching.

Table 3. Social network analysis (SNA) of interactivity in this community

| Network density | Early KB stage | Later KB stage | Whole semester |
|---|---|---|---|
| Note reading | 223 (80.79%) | 223 (80.79%) | 276 (100%) |
| Note linking | 47 (17.02%) | 130 (47.1%) | 143 (51.81%) |
| Build-on | 35 (12.68%) | 109 (39.49%) | 122 (44.2%) |
| Reference | 16 (5.79%) | 15 (5.43%) | 30(10.86%) |

| | | | |
|---|---|---|---|
| Annotation | 17 (6.15%) | 57 (20.65%) | 71(25.72%) |

### Reflective patterns

Table 4 shows how the focus of students' discourse with regard to teachers' cognitive activity/responsibility changed over time from the early to the later knowledge building stages. A Chi-Square test showed a significant difference between the two stages ($\chi^2 = 19.78$, df = 1, $p < .001$). As it shows, at the early KB stage before the midterm, students' online discourse tended to focus more on lower-level cognitive responsibility of teachers, highlighting that teachers only need to understand and appreciate teaching theories, and apply them accordingly in practice. As an example (translated from Chinese), below is a student's online reflection after she read an article about corporal punishment; in her reflection, she basically views theories as authoritative sources for knowledge application:

> The teacher [in the reading materials] expressed her opinions on "education of love" and "corporal punishment". I have no teaching experience, in reality, and therefore am not able to judge which strategy works better. I, however, very much agree with her ideas. "Education of love" certainly could cultivate more independence and autonomy in students, but I also doubt its effectiveness. For those students who appreciate the teacher's encouragement, "education of love" might work well; however, for those who do not care about the teacher's encouragement, "corporal punishment" might not be a bad thing…Nevertheless, I believe that the approach of "education of love" has more benefits than drawbacks…If the teacher can apply the right strategy at the right situation, students will be able to trust him/her.

At the later KB stage, students' online discourse began to focus more on the higher-level cognitive responsibility of teachers, emphasizing that teachers need to improve and even generate their own teaching theories. For example, below is another excerpt from the same student drawn from a note composed at the later knowledge building stage:

> Thank you, those who replied to my note. I am glad to see that we are gradually linking our ideas together. I believe we all think that "a theory needs to be shaped again and again." This is a process and also a procedureof strengthening a theoretical statement. By referring to classmate Hsu's idea, I think theory itself is a conceptual sketch. No matter how it is challenged or shaped by the practice,the sketch will be modified and refined in a better way.

Table 4. Changes in students' understanding of the relationships between theories and practices in teaching

| | Lower-level cognitive activity | Higher-level cognitive activity |
|---|---|---|
| Early knowledge-building stage | 42 | 9 |
| Later knowledge-building stage | 25 | 36 |

The above findings suggest that engaging students in knowledge building practice is helpful to gradually promote more reflective discourse among participants and deepen their understanding of the relationship between theory and practice in teaching. Below we examine whether engaging students in knowledge building practice also has any effects on their views about knowledge building theory and practice.

## Students' views about knowledge building

### Changes in students' perceived importance and feasibility of knowledge building

To further understand if engaging students in knowledge building practice also has impact on their views about the importance and feasbility of knoweldge building, t-tests were conducted. First, in terms of the pre-test, it was found that the teacher-education students tended to consider knowledge building to be both important (M=4.38, SD=0.41) and feasible (M=3.35, SD=0.49) as their means were both higher than the average mean value (M=3.0). To explore futher, however, a paired-sample t-test showed a significant difference between the importance and feasibility of knowledge building (M=1.03; SD=0.68; t=7.41, df=23, p <.01), suggesting a perceived discrepancy among the participants (see Table 5). In terms of the post-test, a paired-sample t-test continued to show that there was a significant perceived discrepancy (M=0.79; SD=0.49) between the importance (M=4.23, SD=0.54) and feasibility (M=3.44, SD=0.44) of knowledge building (t=7.88, df=23, p <.01) at the end of the semester. These findings, however, were quite expected as the participants were teacher-education students who had no prior teaching experience at the time of this study; thus it was natural that they inclined to rate the feasbility lower than the importance, both in terms of pre-test and post-test. Nevertheless, what is more important to know is whether the discrepancy was reduced after engaging students in knoweldge-building practice for a semester. Further t-test indicated there was a marginally significant difference (M=0.24, SD=0.57) between pre-post tests (t=2.02, df=23, p =.055), suggesting that engaging students in knoweldge-building practice did help reduce their perceived discrepancy to some extent.

Table 5. Students' perceived discrepancy between the importance and feasibility of knowledge building

| | Difference | | t-value | p-value |
|---|---|---|---|---|
| | M | SD | | |
| Importance-feasibility discrepancy in pre-test | 1.028 | 0.679 | 7.410 | 0.000** |
| Importance-feasibility discrepancy in post-test | 0.792 | 0.492 | 7.881 | 0.000** |
| Reduced discrepancy between pre-post tests | 0.236 | 0.573 | 2.017 | 0.055* |

* < .10 ** <.01

## Perceived barriers to knowledge building

As the above finding suggests, students' perceived feasibility was relatively low as compared with their perceived importance of knowledge building. With this in mind, a relevant question to ask is what might be the barriers to students perceiving knowledge building as feasible? Making these barriers explict is an essential step to addressing them. Our follow-up interviews indicated concerns regarding the aforementioned three knoweldge building entitites (agency, ideas, and community).

Views on student agency. The interview data first revealed teacher-education students' distrust of children as epistemic agents capable of constructing their own knowledge. For exmaple, as one participant commented, "I think it [knowledge building] is less feasible because of age differences. Age differences must be considered. This is especially true for young students. I believe that if they plan their own learning, they will focus on playing." Apparently, this interviewee tends to believe that children are too young to plan and regulate their own learning as an independent knowledge agent. Such beliefs, however, are contrary to previous research findings that suggest knowledge building is possible even among young children such as grade five students (Hong, Scardamalia, Messina, & Teo, 2008; Zhang, Scardamalia, Lamon, Messina, & Reeve, 2007). Unfortunately, such disbelief in children's knowledge building capacity does align with conventionally held eudcational beliefs which hold that learning must always come first (e.g., during K-12 schooling), before one can really produce new knowledge (e.g., during graduate study) (cf. Hong & Sullivan, 2009). Under this view, maximizing one's individual knowledge (i.e., seeing knowledge as a psychological state confined within Popper's second World) seems an important criterion in judging whether instruction is effective or not, leaving little room for knowledge-building.

Views on idea-centered learning. The interviewed students were also less in favor of idea-centered learning that highlights the importance of sustained production and improvement of ideas. Instead, they tended to emphasize the importance of accumulating basic knowledge in order to pass exams. As an interviewee commented, "…it [knowledge building] is less feasible because what is taught in school in order to help students pass exams is often not related to the real ideas or authentic problems in life." As another commented, "It is not practical to teach more than one solution to a math problem. For example, in learning math, more than one solution [as opposed to idea diversity] may lead students to confusion, especially when the instructional goal is to help students pass the test." As mentioned above, conventional classroom work tends to be defined by pre-specified procedures, clear scripts and rules, and highly-structured learning activities in order to help learners acquire pre-specified knowledge efficiently and then pass exams. As such, establishing a broader knowledge base becomes much more important as an instructional goal than encouraging students to work innovatively with knowledge and engage in sustained idea production and improvement.

Views on community knowledge. The interview data also showed the future teachers' concern about group equity and social harmony in relation to building community knowledge. For example, a participant commented, "I think people can work together in a group but there will never be equal contribution in a group." As another commented, "You can not make sure everyone will have the same value and share the same responsibility, as each one has his or her own individual learning goal." In other words, they still tend to treat knowledge as individual, rather than public, property. They were also less inclined to accept the concept that to give knowledge is to get knowledge in a knoweldge community. For example, one said, "some members never give/share knowledge, but just take from others. To maintain a good social relationship is a key factor that should be taken into consideration." Perhaps, this is because their past schooling and test-related experiences tend to emphasize individual learning rather than group knowledge work. Clearly, how to help transform these teacher-education students' individualistic learning view into a view that also appreciates the social aspects of learning remains an important challenge.

## Summary and Conclusion

In this exploratory research study, we reported the process of knowledge building among a group of teacher-education students and investigated the effects of this knoweldge building process on their views about knowledge building theory and practice. In summary, it was found that engaging students in knowledge building is helpful to (1) promote gradually more interactive and reflective online knowledge building activities; and (2) to somewhat reduce their perceived discrepancy between the importance and feasibility of knowledge building as a theory of knowing and as a way to transform conventional teaching practice. In addition, a major challenge of implementing knowledge building identified through in-depth interviews among participants who especially rated knowledge building as less feasible was that participants' prior schooling experience and socio-cultural expectations tended to strongly influence how they might interpret and value the feasibility of knowledge building. Overall, these participants' prior epistemic views are still largely confined within Popper's world 2 epistemology which sees knoweldge as psychological entity (as opposed to the concept that sees ideas as public artifacts) and learning as individualistic activities (as opposed to the concept that sees learning as a communal acivity) and as accumulation of authorative knowledge (as opposed to the concept that values self-initiated and self-directed knowledge construction). To help students develop more informed and practical views of knowledge building theory and practice thus implies helping them to develop a world 3 knowledge view that sees knowledge as public conceptual artifacts and learning as a social process (Hong, Scardamalia, & Zhang, 2007).

The instructional goal of the present research was (1) to help better prepare teacher-education students to attain a deeper understanding of the relationships between theory and practice in teaching, and (2) to help them develop more informed views about knowledge building. To further this end, we conjecture that a possible strategy is to make teacher-education students' own pedagogical, epistemological, and socio-cultural views about learning and knowledge-building more visible to themselves. Accordingly, an effective instructional design may be to engage them to discuss more explicitly in class their own views about knowledge-building, while at the same time engaging them in actual knowledge-building practices. It is posited that doing so would further help students clarify their conceptual discrepancies between theories and practices in teaching, and gradually achieve World-3 oriented views and thus be able to see knowledge building as more feasible in reality. It is futher conjectured that after being immersed as a knowledge builder in the teacher education program, it may be beneficial to engage teacher-education students in facilitating knowledge-building communities during their practicum experiences under the guidance of experienced knowledge building teachers (See Chai & Tan, 2009). Given the deeply rooted nature of beliefs highlighted above, it seems clear that a single stand-alone course on knowledge building is unlikely to counter the effects of existing beliefs and views on one's own teaching and learning. These claims, however, remains to be further examined by future research.

## References


Anderson, L.W. & Krathwohl, D.R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

Bereiter, C. (1994). Implications of postmodernism for science, or, Science as progressive discourse. *Educational Psychologist, 29*(1), 3-12.

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: LEA.

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Chicago, Illinois: Open Court.

Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. D. Corte, L. Verschaffel, N. Entwistle & J. v. Merrienboer (Eds.), *Unravelling basic components and dimensions of powerful learning environments* (pp. 55-68). Oxford, UK: Elsevier Science.

Burtis, J. (2002). *Analytic Toolkit for Knowledge Forum (Version 4.0).* Toronto, ON: Institute for Knowledge Innovation and Technology, Ontario Institute for Studies in Education/University of Toronto.

Bloom, B.S. (Ed.) (1956). *Taxonomy of educational objectives*. New York: David McKay.

Chai, C. S. & Khine, M. S. (2006). An analysis of interaction and participation patterns in an online learning community. *Journal of Education Technology and Society, 9*(1), 250-261.

Chai, C. S., & Tan S. C. (2009). Professional development of teachers for computer-supported collaborative learning (CSCL) through knowledge building. *Teacher College Records, 111*(5), 1296-1327.

Chai, C. S., Hong, H.-Y., Teo, T. (2009). Singaporean and Taiwanese pre-service teachers' beliefs and their attitude towards ICT use: A comparative study. *Asia-Pacific Educational Researcher, 18*(1), 117-128.

Cohen, D.K. (1989). Teaching practice: Plus que ca change.... In P. W. Jackson (Ed.), *Contributing to Educational Change: Perspectives on Research and Practice*(pp. 27-84). Berkeley, CA: McCutchan.

Creswell, J. W. (2005). *Educational research: planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson.

Hargreaves, D. H. (1999). The knowledge-creating school. *British Journal of Educational Studies, 47*(2), 122-144.

Hong, H.-Y. & Lin, S. P. (2010). Teacher-education students' epistemological belief change through collaborative knowledge building. *Asia-Pacific Educational Researcher, (19)1, 99-110.*

Hong, H.-Y. & Scardamalia, M. (2008). *Using key terms to assess community knowledge.* Paper presented at the annual conference of American Educational Research Association (AERA), New York.

Hong, H.-Y., & Sullivan, F. R. (2009). Towards an idea-centered, principle-based design approach to support learning as knowledge creation. *Educational Technology Research & Development, 57*(5), 613-627.

Hong, H.-Y., Scardamalia, M., & Zhang, J. (2007). *Knowledge Society Network: Toward a dynamic, sustained network for building knowledge.* Paper presented at the annual conference of American Educational Research Association (AERA), Chicago.

Hong, H.-Y., Scardamalia, M., Messina, R., & Teo, C. L. (2008). Principle-based design to foster adaptive use of technology for building community knowledge. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F.J. Prins (Eds.), *International Perspectives in the Learning Sciences: Cre8ing a learning world. Proceedings of the Eighth International Conference for the Learning Sciences – ICLS 2008, Vol. 1* (pp. 374-381). Utrecht, the Netherlands: International Society of the Learning Sciences, Inc.

Hong, H.-Y., Zhang, J., Teo, C. L. & Scardamalia, M. (2009). Towards design-based knowledge-building practices in teaching. In C. O'Malley, D. Suthers, P. Reimann, A. Dimitracopoulou (Eds.), *Computer Supported Collaborative Learning Practices: CSCL2009 Conference Proceedings* (pp. 257-261). Rhodes, Greece: International Society of the Learning Sciences, Inc.

Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2003). Patterns of participation and discourse in elementary students' computer-supported collaborative learning. *Learning and Instruction, 13*, 487-509.

Pajares, M. F. (1992). Teacher's beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach.* London: Oxford Univ. Press.

Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal, 28* (3), 559-586.

Sawyer, K. (Ed.). (2006). *Cambridge handbook of the learning sciences.* Cambridge: Cambridge Univ. Press.

Sawyer, R. K. (2004). Creative teaching. *Educational Researcher, 33*(2), 12-20.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *Journal of the Learning Sciences, 1*, 37-68.

Scardamalia, M. (2003). Knowledge building environments: Extending the limits of the possible in education and knowledge work. In A. DiStefano, K. E. Rudestam & R. Silverman (Eds.), *Encyclopedia of distributed learning* (pp. 269-272). Thousand Oaks, CA: Sage Publications.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage Publications.

van Aalst, J. & Chan, C. K. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *Journal of the Learning Sciences, 16*(2), 175-220.

Wilson, S. M. (1990). The secret garden of teacher education. *Phi Delta Kappan, 72*, 204-209.

Zhang, J., Hong, H.-Y., Teo, C. L., Scardamalia, M., & Morley, E. (2008). *"Constantly Going Deeper:" Knowledge-building Innovation in an Elementary Professional Community.* Paper presented at the annual conference of American Educational Research Association, New York.

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of nine- and ten-year-olds. *Educational Technology Research and Development, 55*(2), 117–145.

## Acknowledgement

The preparation of this paper was supported in part by the National Science Council, Taiwan, grants NSC96-2524-S-008-001 and NSC97-2511-S-004-001-MY2. We extend special thanks to the students and teacher for their participation and for the research opportunities enabled by them. Our special thanks also go to Florence R. Sullivan for editing the manuscript.

# Design-based knowledge building practices in mathematics teaching

Huang-Yao Hong & Yu-Han Chang
National Chengchi University, Taiwan
Email: Hyhong@nccu.edu.tw, 97152010@nccu.edu.tw

**Abstract:** This study investigates the impact of engaging teacher-education students in sustained design and knowledge building practices in mathematics teaching on their views of mathematics and mathematics teaching. Participants were nine teacher-education students who took a university course titled "Middle School Mathematics Teaching." Data primarily came from student discourse recorded in a Knowledge Forum database, peer evaluation and video analysis on student teaching practices, and an open-ended mathematics belief survey. Preliminary results suggest that teacher-education students became progressively more adaptive in their teaching practice. More importantly, their views of mathematics became more constructivist-oriented, and in the meantime, their views of mathematics teaching also became more student-centered.

## Introduction

The purpose of this study is to introduce knowledge building practices into a teacher-education course and to investigate if such practices have any impacts on teacher-education students' views of mathematics and mathematics teaching. According to Paul Ernest (1991), there are two general epistemological stances in terms of views of mathematics, one often referred to as foundationalists and absolutists, the other fallibilists, humanists, relativists and constructivists (see also Handal, 2003). The fundamental argument of the former is that mathematics knowledge is certain, cumulative and unaffected by social interests or personal value. On the other hand, the latter is inclined to believe that mathematics knowledge is through historical, social and cultural, and that there are limitations to its claims of certainty and absoluteness (Ernest, 1991). While the two views are opposing to each other, it is accepted that people's epistemological views in general tend to progress slowly from one that sees knowledge as certain and absolute, to a more relativistic stance that emphasizes knowledge as uncertain and constructed by individuals (Chai, Hong, & Teo, 2009; Magolda, 2004; King & Kitchner, 1994).

As for views of mathematics teaching, they may be broadly classified under the knowledge transmission category or the knowledge construction category (Entwistle, Skinner, Entwistle & Orr, 2000; Handal, 2003; Samuelowicz & Bain, 2001). The former category is characterized as teacher-centered, content-based didactic teaching practice that emphasizes on passive reception of knowledge by students; and the later as student-centered, learning-oriented constructivist instruction that encourages students to actively make sense of their experiences.

Previous research indicates that epistemological views are closely in relation to learning in various ways and have implications for teaching (Pajares, 1992; Richardson, Anders, Tidwell & Lloyd, 1991; Schommer, 1994; Wilson, 1990). Yet, such views are often neglected, or not properly addressed, within teacher education programs (Nespor, 1987). Despite studies in general suggest that it is possible to change pre-service teachers' epistemological views (e.g., see Brownlee, Purdie & Boulton-Lewis, 2001; Gill, Ashton & Algina, 2004; Howard, McGee, Schwartz & Puecell, 2000), there is no consensus as to what represents a most effective way.

One way to help teacher-education students develop more constructivist-oriented epistemological and pedagogical views is perhaps to directly engage them in knowledge building practices (Chai & Tan, 2009; Hargreaves, 1999; Hong, Zhang, Teo, Scardamalia, 2009). Knowledge-building is a social process focused on the production and continual improvement of ideas of value to a community (Bereiter & Scardamalia, 2003) and is supplemented by the use of a software program called Knowledge Forum. Previous research on in-service teachers who have been practicing knowledge building pedagogy for years suggests that it may stimulate epistemological growth among teachers (Chai, Wong & Bopry, 2009; Chai & Tan, 2009; Hong, Zhang, Teo, Scardamalia, 2009). Therefore, it is posited that engaging teacher-education students in collaborative knowledge building practices should also have effects on their views about the subject matter they are to teach and about their teaching methods. Yet, such assumption remains to be examined, especially for the subject of mathematics. The purpose of the present study is to investigate this claim and our main research questions focus on: How do knowledge-building practices affect students' learning processes and outcomes? Specifically, in terms of processes, we looked into participants' online contribution pattern, social interaction patterns, and patterns in their teaching practices throughout the semester; and in terms of outcomes, we looked into changes in their views of mathematics and mathematics teaching.

## Method

### Context and participants

The present research was conducted in a university course titled "Middle School Mathematics Teaching" in Taiwan. The course was offered by the university's Center of Teacher Education to students who plan to become a mathematics teacher at the middle school level. The university is ranked as one of the best universities in the nation and the students enrolled in the subject university are all academically high-achievers and are overall considered by the society as the best prospective teachers in the nation. Participants in this study were four female and five male teacher-education students and their age ranged from 19 to 23 years (M = 21; SD=1.59).

### Instructional approach

We employed an instructional approach that features design-based knowledge building practices (Hong, Zhang, Teo, & Scardamalia, 2009). The main purpose of engaging teacher-education students in such practices is threefold: (1) to help them understand the nature of teaching as creative and adaptive (Sawyer, 2004), rather than routine and script-oriented, (2) to help them avoid viewing teaching as merely pursuing best practices of certain model teachers by means of mastering pre-defined teaching skills; and (3) to enable them to go beyond "best practices" and assume the role of designer and knowledge-builder in continual improvement of their own teaching practices.

To implement, the course was divided into the following four related phases: (1) Initial Design: First, participants were guided to work on their initial design in order to implement their first teaching practice; accordingly, they prepared lesson plans, and learning materials and learning sheet, etc.; (2) First teaching practice: Then, based on their initial design, they performed their first teaching practices in class, with the other classmates serving as the audience. The whole process of each student's teaching practice was video-taped, and, at the end of each teaching practice, a peer evaluation was conducted to assess the quality of teaching for future improvement; (3) Re-design: During this phase, participants collectively worked online in Knowledge Forum to provide feedback and suggestions to the target student who already performed his (or her) teaching practice. Then he (or she) further reflected on these suggestions, analyzed the recorded video of his (or her) own teaching practice, and collaborated with other peers to improve his (or her) initial design by co-designing next lesson; and (4) Second teaching practice: Finally, based on the new instructional design, each participant performed their second teaching practice; the whole teaching process again was video-taped, and then a second peer evaluation on the quality of teaching was conducted again.

In addition, the course also employed other complementary instructional activities, mainly whole-class and small-group discussion after each teacher-education student's teaching practice, in order to engage them to frequently reflect on how to improve their own and others' teaching practices. Some questions being discussed in class were, for example: what have you learned from others' teaching practices; if you were to teach this lesson again, what would you do differently, and why? There were no fixed or pre-defined answers to most questions. The instructor served as a facilitator in guiding students to explore, discuss and reflect on all questions emerged in class in order to help them construct their own views of mathematics and mathematics teaching. Knowledge Forum were only used after class and it played an important role as an online space for students to record all key points generated from class discussion and personal reflection.

### Knowledge Forum environment

The learning environments adopted in the present study is Knowledge Forum (KF). It is an online platform that runs on a multimedia database. KF allows users to simultaneously create and post their ideas in the form of note into the database, read others' postings, reply to others' notes, search and retrieve records, and organize notes into more complex knowledge representation. KF runs on both a text-based and graphics-based mode. In the graphics mode, it shows linkages of postings as a way to represent the interconnectivity and dialogical nature of knowledge. As such, it also enables the development of ideas to be traced. Figure 1 shows an example of a KF view (i.e., an open space designed for collaborative problem-solving), within which users are guided to work as a community by posting their problem of interest, producing initial ideas for problem-solving, sharing and connecting ideas, synthesizing their ideas, and deepening their collective understanding of problems at issue. Specifically for this study, the main problem of interest is concerned with understanding of the nature of mathematics and mathematics teaching, and in the beginning of the course students generated and shared their best ideas (e.g., "What is mathematics?"); then they designed and practiced their mathematics teaching according to their initial ideas and try to improve their initial ideas about mathematics teaching through frequent discussion, reflection, and co-design in Knowledge Forum.

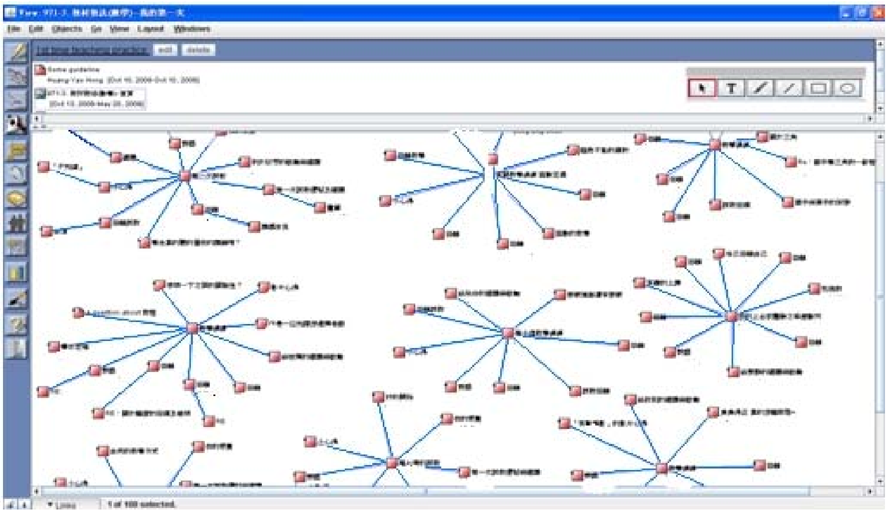


Figure 1. A Knowledge Forum view

## Data Source and Analysis

Data source mainly came from student online discourse, two teaching practices, and a belief survey. First, in terms of student online discourse, it was recorded in a Knowledge Forum database and we analyzed student participation patterns (e.g., note creation and reading) and interaction patterns by employing descriptive analysis and social network analysis. In addition, we also looked into patterns of group feedback and personal reflection, in terms of each participant's own teaching practices.

The second set of data came from peer evaluation and video analysis on students' two teaching practices. In the beginning of the course, students collectively constructed an evaluation form focusing on the following five categories: objective (one item), learning materials (five items), instructional methods (six items), activities (six items), and presentation (three items). All items employed a 5-point Likert scale (5=excellent performance, 1=poor performance). In addition, both teaching practices of each participant were videotaped and then content-analyzed based on a pre-determined coding scheme highlighting the following three general types of learning activity: passive, active, and interactive (see Collins, 1996). To analyze, we examined percentage of time spent in each type of activity during each teaching practice.

The third set of data came from a belief survey, which was designed based on Handal's (2003) conceptualization of mathematics beliefs concerning the nature of mathematics and that of mathematics teaching. Accordingly, we asked the following eight open-ended questions: (1) What is mathematics? (2) What does "doing mathematics" mean? (3) What is an ideal way to teach mathematics? (4) What are some key factors for successful mathematics teaching? (5) What makes an ideal mathematics teacher? (6) What is an ideal way to learn mathematics? (7) What are some key factors for successful mathematics learning? (8) What does an ideal mathematics learning environment mean to you? Of all questions, questions 1 and 2 concern the nature of mathematics; and questions 3 to 8 concern the nature of mathematics instruction. To analyze, we employed an open coding procedure (Strauss & Corbin, 1990). Six codes emerged from this coding process (see table 1). Pair-sampled t-tests were conducted to examine if there were any pre-post differences.

Table 1: Coding scheme for students' views of mathematics and mathematics teaching

| Categories | Code | Example of students' statements (Translated from Chinese) |
|---|---|---|
| Nature of mathematics | A science of quantity and calculation applied in life | Math is a science about calculating numbers.<br>What we see, do and make in our daily life all needs math. |
| | A science of patterns and orders | To do mathematics is to seek for the patterns or rules by means of certain given conditions, using symbols and numbers to predict, estimate, or conjecture possible outcomes. |
| Nature of mathematics teaching | To let students practice again and again | Practice makes perfect. |
| | To let students understand basic concepts and procedures | To understand what each mathematical formulas means and to get the gist of a math problem and know how to solve it. |
| | To train logical thinking skills | The most important thing is to train students' problem-solving and thinking abilities. |

| | | |
|---|---|---|
| | To help students develop and explore their own way of mathematics learning | Students have to explore and find their own way of math learning and develop their own learning styles. |

## Results and discussion

### Online contribution patterns

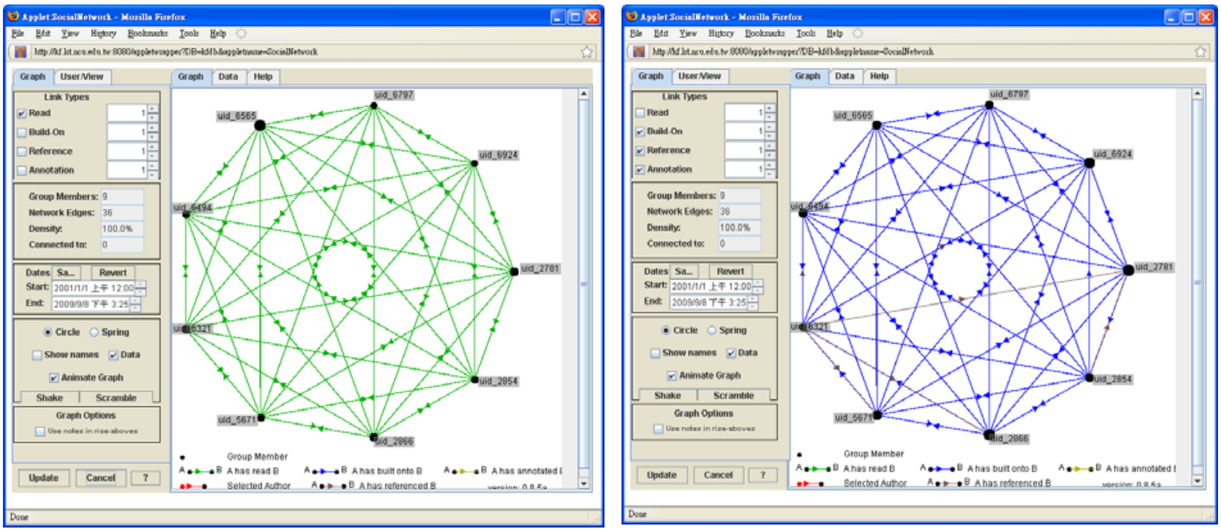
The overall online activity and performance in this community is shown in Table 2. Throughout the whole semester, participants contributed a total number of 171 notes with a mean number of 17.8 (SD=4.29) notes per person. In addition, other complementary online knowledge-building measures recorded in this community include number of note read, percentage of note read, number of annotation, number of note revisions, number of build-on notes, and percentage of notes linked. Overall, the online activities were substantive as compared with previous research using subjects with similar background (Hong & Lin, accepted). Nevertheless, while these behavioral measures gave a general picture of how participants worked online in this database, they do not tell much about how participants actually interacted with one another. To better understand the social dynamics in the community, a social network analysis (SNA) focusing on network density was conducted.

Table 2. Descriptive analysis on individual online knowledge-building activities

| | Mean | SD |
|---|---|---|
| No. of notes created | 17.8 | 4.29 |
| No. of notes read | 140.2 | 32.94 |
| Percentage of notes read | 82% | 19.26% |
| No. of annotations | 21.2 | 12.26 |
| No. of note revisions | 8.2 | 3.29 |
| No. of build-on notes created | 11.3 | 2.49 |
| Percentage of notes linked | 64.3% | 6.17% |

SNA was conducted using the automatic assessment tools embedded in the Knowledge Forum. Figure 2 shows the overall interactive and collaborative patterns in the community throughout the whole semester, using two indicators that are available in the Knowledge Forum: "note reading" (which indicates community awareness of contributions made by other peers) and "note building-on" (which indicates complementary contributions by the effort to build on to others' work and ideas). Table 3 shows detailed results of participants' interactions in two knowledge-building stages (using mid-term exam as a point of separation). In this particular analysis, we used an indicator, called "network density", which is defined as the proportion of connections in a network relative to the total number possible. The higher the number of the density is, the stronger the social dynamics a community is implied. An intention of adopting knowledge-building practices in this course is to transform the traditional way of knowledge-transmission mode of learning into a knowledge-construction mode that engages these students in collective problem-solving and knowledge work. Therefore, it was expected that students should progressively work more collaboratively in Knowledge Forum. As expected, there was an increasing trend of social interactions as reflected by the measures of density recorded online for this community from the early to the later knowledge building stages. Lipponen Rahikainen, Lallimo, and Hakkarainen (2003) regarded a social network density of .39 for students building-on each other online messages as adequate. In the present study, the density level is at .94, which indicates a highly satisfactory social dynamics of this community. The SNA findings alone, however, did not tell us much about the quality of interaction in the community. So, we further content-analyzed students' notes to illustrate what they actually did to help each other improve their teaching practices.

Table 4 shows the total numbers of group feedback and personal reflection made after the first and before the second teaching practice in terms of three dimensions: instructional design, learning materials, and presentation skills. There were in total 106 suggestions/comments and 43 times of personal reflections being made. On average each student received 13.25 suggestions from others and made 4.78 times of personal reflection between the first and the second teaching practices. This suggests that participants' online interaction were both quite purposeful and practical. The next question to ask is how online interaction, group feedback, and personal reflection contribute to the improvement of student actual teaching practices.



a. Note reading b. Note linking

Figure 2. Interaction patterns in the community throughout the whole semester

Table 3. Social network analysis (SNA) of interactivity in the community

| Network density | Early KB stage | Later KB stage | Whole semester |
|---|---|---|---|
| Note reading | 100% | 100% | 100% |
| Note building-on | 72.22% | 94.44% | 100% |

Table 4. Group feedback and personal reflection made between the first and the second teaching practices

| Category | Dimension | Frequency |
|---|---|---|
| Group feedback | 1. Instructional design | 44 |
| | 2. Learning materials | 28 |
| | 3. Presentation skills | 34 |
| Personal Reflection | 1. Instructional design | 16 |
| | 2. Learning materials | 14 |
| | 3. Presentation skills | 13 |

## Patterns in teaching practices

Table 5 shows students' peer evaluation results. It was found that the ratings in the second teaching practice were significant higher than the first teaching practice in 20 (out of the total 24) aspects of teaching practice. This suggests that there were significant improvement in students' second teaching practice. While the findings showed no significant improvement in four aspects—"Clarity of instructional goal" (M=4.49), "Appropriateness of learning materials" (M=3.98), "knowledge of the learning content" (M=4.38), and "Time management" (M=4.19)—this could be due to ceiling effects as their initial ratings were already very high (i.e., closer to or beyond point 4, out of the maximum points of five in a Likert scale). One thing worth noting is that the biggest improvement mainly came from the "Instructional methods" category which in particular highlights adaptive teaching (e.g., Uses innovative instructional approaches) and student-centered learning (e.g., Promots creative, critical and higher-level thinking), rather than conventional step-by-step and scripted teaching methods. These changes are important indicators of adaptive teaching as a central goal of employing design-based knowledge building practices in the present study was to help participants achieve a more flexible and innovative teaching approaches.

Table 5. Peer evaluation on one another's teaching practices

| Category | Aspect | 1st practice | | 2nd practice | | t-value |
|---|---|---|---|---|---|---|
| | | M | SD | M | SD | |
| Objective | Establishes clear purpose or objective of lesson | 4.49 | 0.13 | 4.58 | 0.15 | -1.31 |
| Materials | Materials are purposeful and interesting | 3.72 | 0.32 | 3.88 | 0.29 | -3.04* |
| | Materials are age-appropriate | 3.98 | 0.32 | 4.22 | 0.16 | -2.29 |
| | Materials are integrally designed and developed | 4.09 | 0.26 | 4.25 | 0.15 | -2.60* |

5

| | | | | | | |
|---|---|---|---|---|---|---|
| | Materials are ready and available | 4.31 | 0.35 | 4.52 | 0.18 | -2.64* |
| | Materials are rich | 4.09 | 0.23 | 4.33 | 0.21 | -3.26* |
| | Knows the content of the subject very well | 4.38 | 0.26 | 4.52 | 0.18 | -2.05 |
| | Uses visual aids (handouts, manipulatives, pictures, etc.) | 3.57 | 0.15 | 4.01 | 0.31 | -6.41*** |
| Methods | Uses innovative instructional approaches | 3.56 | 0.47 | 3.88 | 0.42 | -4.56*** |
| | Maintains academic focus | 3.67 | 0.43 | 4.04 | 0.32 | -4.47*** |
| | Begins lesson with attention getter | 3.64 | 0.51 | 4.11 | 0.39 | -6.19*** |
| | Maintains high percentage of student involvement | 3.81 | 0.46 | 4.42 | 0.26 | -5.44*** |
| | Promotes creative, critical and higher-level thinking | 3.59 | 0.40 | 3.96 | 0.38 | -7.07*** |
| | Instruction is adaptively designed | 3.48 | 0.47 | 3.65 | 0.45 | -3.28* |
| Activities | Manages time well | 4.12 | 0.17 | 4.22 | 0.12 | -2.16 |
| | Applies established rules for behavior consistently | 4.19 | 0.22 | 4.32 | 0.15 | -2.53* |
| | Maintains instructional momentum | 4.32 | 0.17 | 4.38 | 0.16 | -2.35* |
| | Maintains an enjoyable learning environment | 3.67 | 0.14 | 3.83 | 0.16 | -3.16* |
| | Utilizes appropriate assessment techniques and practices | 3.72 | 0.44 | 3.87 | 0.47 | -4.26** |
| | Presents activities appropriate for all students | 4.12 | 0.17 | 4.22 | 0.12 | -3.42** |
| Presentation | Exhibits positive body language related to content | 4.32 | 0.46 | 4.47 | 0.35 | -3.02* |
| | Appropriately varies volume and inflection | 4.10 | 0.39 | 4.36 | 0.28 | -4.67** |
| | Uses clear, unscrambled discourse | 3.86 | 0.52 | 4.15 | 0.25 | -2.55* |
| | Maintains eye contact with everyone | 4.04 | 0.36 | 4.28 | 0.19 | -3.59** |

* p<.05 ** p<.01 *** p<.001

Video analysis was further conducted as a form of data triangulation. Table 6 shows analysis results in terms of percentage of time spent in three different instructional activities during the two teaching practices. It was found that there was a significant decrease in the percentage of time spent in passive learning activities, from the first practice (72.1%) to the second practice (46.9%) (t=5.04, df=8, p<.01). In contrast, there was a significant increase in the percentage of time spent in active learning activities, from the first practice (17.9%) to the second practice (36.4%) (t=-3.79, df=8, p<.01), and a slight increase in the percentage of time spent in interactive learning activities (t=-2.15, df=8, p=.064). Overall, our video analysis confirmed that participants' teaching practice was shifting from a more teacher-centered approach to a more student-centered approach.

Table 6. Percentage of time spent in different instructional activities in two teaching practices

| Activity | First practice | Second practice |
|---|---|---|
| Passive learning | 72.1% | 46.9% |
| Active learning | 17.9% | 36.4% |
| Interactive learning | 10.0% | 16.7% |
| Total | 100.0% | 100.0% |

## Views of mathematics and mathematics teaching

Finally, we looked into whether engaging students in design-based knowledge building practices also have effects on their views of mathematics and mathematics teaching. As shown in Table 7, in terms of views of mathematics, there was a significant drop in scores in the view that emphasizes mathematics as a science of quantity and calculation (M = 3.67 in the pre-survey and M = 1.67 in the post- survey). In contrast, there is a significant increase in scores in the view that highlights mathematics as a science of patterns and orders between the two measurements (M = 0 in the pre-measurement and M = 2.33 in the post- measurement). In sum, the findings suggest that in the beginning of the semester, teacher-education students tended to think that mathematics was all about number calculation or was something that is useful and can be applied to our daily life. Accordingly, it was assumed that a teacher should focus his or her instructional goal on helping students do appropriate mathematics calculation or develop related math abilities. However, such routine mathematics practices and training may only help students be familiar with and able to acquire some conceptual and procedural mathematics knowledge efficiently. It may not necessarily help students to use mathematics in a more creative way. But, after a semester, it was found that the teacher-education students were able to gradually view mathematics from a more constructive and creative manner. They began to appreciate mathematics as a science of finding patterns and orders, or as a tool for more constructive learning and problem-solving.

Accordingly, they also changed their views of mathematics teaching. As shown in the bottom part of

6

Table 7, there was a desirable (although insignificant) drop in scores in terms of the first three views of mathematics teaching, which regard mathematics teaching as a way "to let students practice again and again" (M = 2.11 in the pre-test and M = 0.44 in the post-test), or a way "to let students understand basic concepts and procedures" (M = 2.56 for the pre-test and M = 1.33 for the post-test), or a way "to train students' logical thinking ability" ( M = 4 for the pre-test and M = 3.11 for the post-test). In contrast, there is a significant increase in scores between the pre-measurement (M = 0.22) and the post-measurement (M = 3.56) in terms of the last view of mathematics teaching, which highlights mathematics teaching as a way "to help students develop and explore their own way of math learning". Overall, it was found that in the beginning of the semester, teacher-education students' views of mathematics teaching is more teacher-led, whereas after engaging in design-based knowledge building practices for a semester, their views of mathematics became more student-centered.

Table 7: Students' views on the nature of mathematics and that of math teaching

| Views of mathematics and mathematics teaching | Pre-survey | | Post-survey | | *t-value* |
|---|---|---|---|---|---|
| | M | SD | M | SD | |
| Nature of Mathematics | | | | | |
| A science of quantity and calculation applied in life | 3.67 | 1.87 | 1.67 | 1.41 | 4.00** |
| A science of patterns and orders | 0 | 0 | 2.33 | 1.80 | -3.88** |
| Nature of mathematics teaching | | | | | |
| To let students practice again and again | 2.11 | 2.21 | 0.44 | 0.73 | 2.041 |
| To let students understand basic concepts and procedures | 2.56 | 3.05 | 1.33 | 1.80 | 0.854 |
| To train students' logical thinking ability | 4.00 | 2.50 | 3.11 | 3.52 | 0.567 |
| To help students develop/explore their own way of learning | 0.22 | 0.44 | 3.56 | 2.19 | -4.588** |

**p<.01

## Summary and implication

In this study, we reported an instructional approach of design-based knowledge building practices among a group of teacher-education students and investigated the effects of such instructional approach on their views of mathematics and mathematics teaching. The instructional approach was design-based as participants were engaged in sustained design, re-design, and co-design activities when planning and practicing their mathematics teaching. The instructional approach was knowledge building oriented as it highlighted continual production and improvement of ideas in pursuit of deeper understanding of the nature of mathematics and mathematics teaching. In summary, it was found that design-based knowledge building practices as an instructional approach was helpful to promote more interactive and reflective online activities. Moreover, it was found that at the end of the course, teacher-education students changed their views of mathematics to become more relativist-oriented and they also changed their views of mathematics teaching to become more student-centered. Arguably, while cultivating teachers' pedagogical content knowledge is important and should always be included as part of overall teaching training in a teacher-education program, our study suggests that it is equally important to help teacher-education students develop more informed and sophisticated mathematical beliefs. And to enable such belief change, it will be crucial for teacher education programs to avoid traditional didactic ways of teacher training, and adopt more constructivist-oriented instructional approaches in order to cultivate more future teachers who view mathematics teaching as creative and adaptive, rather than routine and ritualistic, practices.

## References

Chai, C, S., Wong, B., & Bopry, J. (2009). A Case Study of Teachers' Perceptions of the Knowledge Building Community. *Asian Journal of Educational Research and Synergy, 1*(1), 69-81.

Chai, C. S., & Tan S. C. (2009). Professional development of teachers for computer-supported collaborative learning (CSCL) through knowledge building. *Teacher College Records, 111*(5), 1296-1327.

Chai, C. S., Hong, H.-Y., Teo, T. (2009). Singaporean and Taiwanese pre-service teachers' beliefs and their attitude towards ICT use: A comparative study. *Asia-Pacific Educational Researcher, 18*(1), 117-128.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. Wittrock (Ed.),*Handbook of research in teaching* (pp. 255-296). New York: MacMillan.

Collins, A. (1996). Design issues for learning environments. In Vosniadou, S., Corte, E. E., Glaser, R. & Mandl, H. (Eds.). *International perspectives on the design of technology-supported learning environments.* (pp. 347-361). Hisdale, NJ: Lawrence Erlbaum Associates, Inc.

Drucker, P. F. (1986),*Innovation and entrepreneurship: Practice and principles.* NY: Harper Business

Entwistle, N., Skinner, D., Entwistle, D., & Orr, S. (2000). Conceptions and beliefs about "good teaching": An integration of contrasting research areas. *Higher Education Research & Development*, *19* (1), 5-26.

Ernest, P. (1991). *The philosophy of mathematics education*. London: The Falmer Press.
Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research,38*(1), 47-65.
Handal, B. (2003). Teachers' mathematical beliefs: A review. *The Mathematics Educator,13*(2), 47-57.
Hong, H.-Y. & Lin, S. P. (2010). Teacher-education students' epistemological belief change through collaborative knowledge building. *The Asia-Pacific Educational Researcher, (19)1, 99-110*.
Hong, H.-Y., & Sullivan, F. R. (2009). Towards an idea-centered, principle-based design approach to support learning as knowledge creation. *Educational Technology Research & Development, 57*(5), 613-627.
Hong, H.-Y., Scardamalia, M., Messina, R., & Teo, C. L. (2008). Principle-based design to foster adaptive use of technology for building community knowledge. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F.J. Prins (Eds.), *International Perspectives in the Learning Sciences: Cre8ing a learning world. Proceedings of the Eighth International Conference for the Learning Sciences–ICLS 2008, Vol. 1* (pp. 374-381). Utrecht, the Netherlands: International Society of the Learning Sciences, Inc.
Hong, H.-Y., Zhang, J., Teo, C. L. & Scardamalia, M. (2009). Towards design-based knowledge-building practices in teaching. In C. O'Malley, D. Suthers, P. Reimann, A. Dimitracopoulou (Eds.), *Computer Supported Collaborative Learning Practices: CSCL2009 Conference Proceedings* (pp. 257-261). Rhodes, Greece: International Society of the Learning Sciences, Inc.
Kagan, D. M. (1992). Implications of research on teacher belief. *Educational Psychologist, 27,* 65-90.
King, P. M., & Kitchener, K. S. (1994). Developing reflective judgment: *Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass.
Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2003). Patterns of participation and discourse in elementary students' computer-supported collaborative learning. *Learning and Instruction, 13*, 487-509.
Magolda, M. B. (2004). Evolution of a constructivist conceptualization of epistemological reflection. *Educational Psychologist, 39*(1). 31-43.
Nespor, J. K. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies,19*(4), 317-328.
Pajares, M. F. (1992). Teacher's beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332.
Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal, 28* (3), 559-586.
Sawyer, R. K. (2004). Creative teaching: collaborative discussion as disciplined improvisation. *Educational Researcher, 33*(2), 12-20.
Samuelowicz, K. & Bain, J. D. (2001). Revisiting academics' beliefs about teaching and learning. *Higher Education*, 41, 299-325.
Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.
Scardamalia, M. (2004). CSILE/Knowledge Forum®. In *Education and technology: An encyclopedia* (pp. 183-192). Santa Barbara: ABC-CLIO.
Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In *Encyclopedia of Education* (2nd ed., pp. 1370-1373). New York: Macmillan Reference, USA.
Schommer, M. (1994). Synthesizing epistemological belief research: Tentative understandings and provocative conclusions. *Educational Psychology Review*, 6, 293 -319
Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: grounded theory procedures and techniques* Newbury Park, CA: Sage Publications.
Van De Walle. A. J. (2001). *Elementary & Middle School Mathematics*: Teaching developmentally, NY: Longman.
Wilson, S. M. (1990). The secret garden of teacher education. *Phi Delta Kappan, 72*, 204-209.
Zhang, J., Hong, H.-Y., Teo, C. L., Scardamalia, M., & Morley, E. (2008). *"Constantly Going Deeper": Knowledge Building Innovation in an Elementary Professional Community.* Paper presented at the annual conference of American Educational Research Association (AERA), New York.

## Acknowledgements

The preparation of this paper was supported in part by the National Science Council, Taiwan, grant NSC97-2511-S-004-001-MY2, and the National Chengchi University's Top University Program grant.

# Centering a Professional Learning Community on a Learning Progression for Natural Selection: Transforming Community, Language, and Instructional Practice

Erin Marie Furtak, Deborah Morrison, and Kathleen Henson
University of Colorado at Boulder, School of Education, UCB 249, Boulder, CO 80309
erin.furtak@colorado.edu, deb.morrison@colorado.edu, kathleen.henson@colorado.edu

In this study we examine the development of a professional learning community (PLC) centered on the practice of understanding student learning of natural selection. Eight biology teachers engaged with researchers to use an educative learning progression (LP) empirically developed from student work as a tool for exploring prior ideas and sharing potential instructional strategies to move students towards scientific understandings of natural selection. The PLC involved iterative cycles of development and reflection of the LP, formative assessments and feedback strategies. Three primary results are reported in this paper with respect to the PLC's development associated with the LP: transformation of interactions within the PLC, changes in the language teachers used to describe student thinking, and modifications in the ways instructional practices were shared. These changes demonstrate the potential for a PLC when used in conjunction with a LP to encourage meaningful changes in classroom practice.

## Theoretical Background

The *National Science Education Standards [NSES]* (National Research Council, 1996) defined the knowledge and practices students are expected to learn, but were never intended as a 'how-to' document for teachers (Collins, 1998). To supplement the NSES, science education researchers are currently creating Learning Progressions (LPs), which are empirically-based sequences of conceptual understandings and scientific skills intended as guides for teaching and learning. LPs map "the successively more sophisticated ways of thinking about a topic that can follow and build on one another as children learn about and investigate a topic over a broad span of time" (National Research Council, 2007, p. 213).

At least two approaches have emerged for the construction of LPs. The first approach consists of a progression of *correct* ideas based on standards documents or consultations with scientists, articulated across grade spans (American Association for the Advancement of Science, 2001; Catley, Lehrer, & Reiser, 2005). This kind of LP tends to be created by exhaustive conceptual analyses of how scientifically accurate student ideas should develop during instruction; examples have been created in the domain of evolution (Catley et al., 2005) and modern genetics (Duncan, Rogat, & Yarden, 2009). These LPs are being developed as guides for curriculum and large-scale assessments. A second approach to LPs entails the creation of maps of concepts anchored on one end by naïve ideas about the natural world and on the other end by scientifically accepted explanations. The middle is occupied by increasingly sophisticated intermediate understandings. A key feature of this second type of learning progression is that they are anchored at the lower end by students' common prior ideas and misunderstandings (Briggs, Alonzo, Schwab, & Wilson, 2006; Wilson, 2005).

### Toward Learning Progressions as Teacher Development Tools

Given the aforementioned challenges presented by previous studies employing LPs in the classroom, it follows that LPs could be complimented with additional information to make them not only maps of how student knowledge develops through instruction, but also maps of the instructional strategies that will help teachers more effectively teach a given concept. Knowing the means by which students learn concepts is important, but without the tools to help teachers recognize and act upon those ideas, the means by which the information LPs contain will be translated into practice is unclear. LPs need to be developed that will scaffold the development of the knowledge and practices that teachers need for effective science instruction.

The science education community, after several years of developing LPs, is now emphasizing the need for empirical explorations of the impact of LPs on teacher practice and student learning (Shavelson, 2009). According to the Center on Continuous Instructional Improvement's recent report on LPs, the field needs "continuing effort, and funding, to refine and test the progressions that now are under development, including making a concerted effort to demonstrate the effects on instruction and improved student outcomes when teaching and assessment are aligned with well tested progressions - to provide a kind of "existence proof" that further investment in developing progressions would be justified." (Corcoran et al., 2009, p. 9).

## Educative Learning Progressions

Although the primary purpose of LPs as currently conceived is to sequentially map the knowledge and practices students are expected to learn, LPs have great potential to be invaluable tools for teacher development. LPs could serve as the foundation for what Davis & Krajcik (2005) called 'educative curriculum materials' with the dual potential of scaffolding not only student learning, but also teacher learning in a conceptual domain. This kind of LP – which will henceforth be referred to as an educative LP - contains information regarding knowledge of common student ideas and alternative conceptions, suggestions for strategies or actions to help students learn, and a model for teacher development in relation to the LP. An educative LP is thus a suite of tools that may work synergistically to support improvement in teacher practice and student learning.

In the inquiry-based learning environments currently emphasized in science education reforms (National Research Council, 1996, 2001, 2007) the ideas that students develop are often undifferentiated, poorly stated, and inconsistent with standards-style statements of idealized student understanding (e.g. Driver, Guesne, & Tiberghien, 1985). In order for teachers to navigate these ideas and move students toward learning goals, they must have deep content knowledge, as well as knowledge of students' understandings and misconceptions (Duschl, 2003; Gess-Newsome, 1999; Shulman, 1986). Since teachers' content knowledge is not always interconnected or easily translated into reform-based classroom practice, the ideas students present in inquiry-based activities pose a particular challenge (Luft & Roehrig, 2007; Windschitl, 2004).

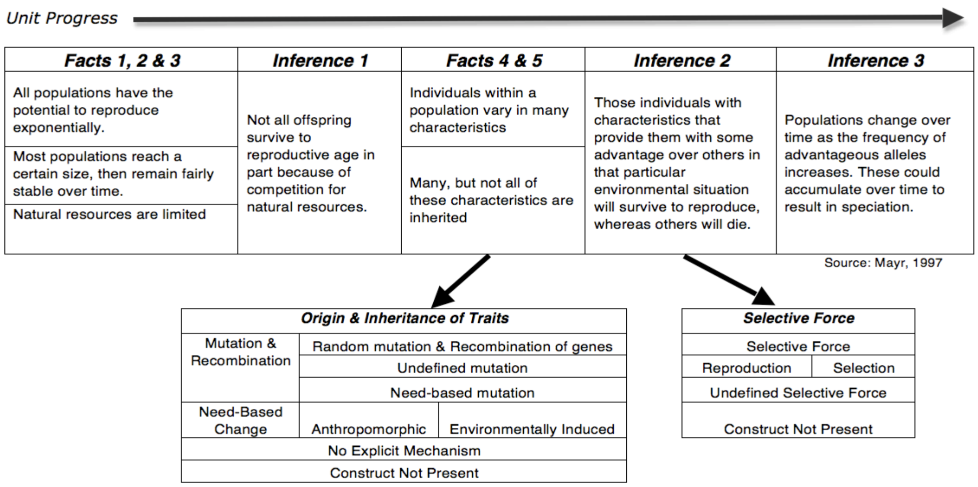
Educative LPs could serve as scaffolds for teachers in inquiry-based settings by helping them to anticipate not only the ideas students might share, but also to suggest instructional strategies tailored to these ideas. In order to serve that purpose, however, LPs need to include examples of student responses for each level, as well as instructional strategies, feedback suggestions to help students proceed up the LP, and professional development to support changes in practice. Educative LPs should thus contain the preceding five elements; the first four are features of the educative LP itself, and the fifth emphasizes the importance of sustained professional development within a community of practice to support teachers' adoption of new instructional strategies based on the educative LP (Lave & Wenger, 1991). Sustained professional development centered on an educative LP and its accompanying tools could help teachers learn the structure and relationship of students' understandings about a particular concept, teaching strategies to elicit student thinking, likely student responses, and targeted feedback for students. In this way, educative LPs can act as a guide for teachers, showing them potential steps to take in order to help students move along the learning progression, thereby meeting increasingly challenging learning goals (Sadler, 1989) and positively impacting student learning (Black, 1998). To explore the potential influence of an educative LP on a PLC, this paper responds to the following research question: How does sustained, ongoing professional development in a PLC centered on the LP transform the community, language and practice?

## Context: Natural Selection and the Daphne Project

Charles Darwin's theory of evolution through the process of natural selection is the unifying framework for Biology (Dobzhansky, 1973); nevertheless, students often have a difficult time understanding it (Bishop & Anderson, 1990; Ferrari & Chi, 1998). Studies indicate evolution instruction in high school has been "absent, cursory, or fraught with misinformation" (Rutledge & Mitchell, 2002, p. 21), and students commonly have misconceptions about how populations change over time (Anderson, Fisher, & Norman, 2002; Rudolph & Stewart, 1998; Shtulman, 2006).

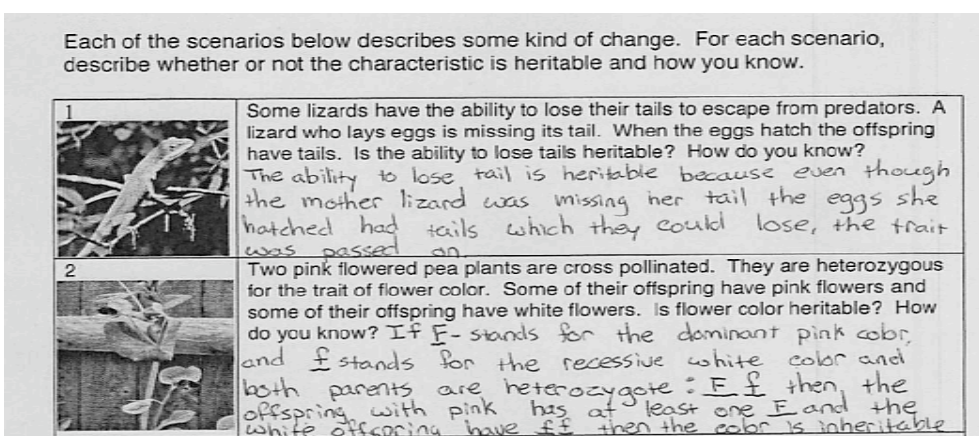
The Daphne Project is a design-based research study located at the intersection of the new theory of LPs and classroom practice (Brown, 1992; Sandoval & Bell, 2004; Sloan 2009). The goal of the Daphne Project was to collaborate with a department of biology teachers to explore and refine a hypothesized LP. It was hypothesized at the outset of the study that engaging in the process of developing this LP would help teachers be better prepared to recognize and act upon student thinking, which in turn would impact student learning by closing what has been called a 'feedback loop' in the formative assessment process (Black & Wiliam, 1998; National Research Council, 2001).

The educative LP created in the Daphne Project combines features of LPs designed as frameworks for curriculum with lower anchors serving as foundational *correct* understandings (Catley et al., 2005) and those anchored by students' common prior ideas as suggested by Shepard (2009) and Wilson (2009). In this way, the LP has as its foundation both an overarching framework to inform scope and sequence as well as information mapping students' complex thinking that can often obfuscate the path forward for teachers. The educative LP is not a curriculum in itself, but is intended as a framework to guide teachers' enactment of a curriculum by representing a sequence of correct ideas as well as a categorization of the common misunderstandings students may have regarding those ideas. The educative LP is shown in Figure 1; the horizontal sequence of correct ideas is taken directly from Mayr (1997), and the two vertical sequences (*Origin and Inheritance of Traits* and *Selective Force* constructs) were identified through analyses of student responses to the assessments in the Daphne Project. The educative LP is paired with accompanying tools intended to scaffold instruction, as well as a professional development model to support teachers in changing their practice. The educative LP for natural selection includes a set of formative assessment prompts designed to be embedded in the curriculum to provide opportunities for teachers to elicit student thinking and map it to the progression. It also includes common student responses to each assessment at each level and suggested feedback strategies for students with different ideas.




<u>Figure 1.</u> Learning Progression for Natural Selection

A suite of tools including a series of formative assessment prompts, sample student responses, and feedback ideas were developed and piloted across the three years of the Daphne Project. The result was a set of formative assessments that targeted different ideas on the *Origin of New Traits* and *Selective Force* constructs, including a constructed-response prompt of "How do species change over time?", a multiple-choice plus-justification prompt, an evaluation of Peter and Rosemary Grant's data on Darwin's finches in the Galapagos Islands, and interpretation of real-life scenarios (Figure 2). and analysis of these prompts being implemented led to collection of student responses for each level of the LP, as well as suggested feedback strategies (Table 1).




*<u>Figure 2</u>*. Sample formative assessment activity

*Table 1*. Sample student responses, and feedback strategies linked to the need-based conceptions on the *Origin and Inheritance of Traits* construct.

| Description | Example Student Responses | Suggested Feedback |
|---|---|---|
| *Anthropomorphic*: Changes arise from deliberate acts by the parent organism or the species as a whole | "The moths become darker because of bark." "The moths would change their color to a darker color to blend in with the dark bark." | Examine evidence collected in class activities and focus upon where variation originated. Explore Lamarck's idea of descent with modification, and compare to modern ideas about genetic inheritance |
| *Environmentally Induced*: Changes emerge because of environmental conditions; i.e., the environment causes changes in organisms to occur | "When food and other resources become scarce, the finches develop over a period of time different beaks to eat the hard seeds that seem to be abundant." "Animals mutate to fit in with their natural surroundings. So becoming darker helps to keep them in camouflage." | Explore differences between mutations in an individual and population changes. Discuss instances in which individuals change in response to the environment (e.g. arctic hares), and how this differs from changes in the proportion of genes in a population |

### Sustained Professional Development

Since the LP and the tools themselves do not represent a sufficient intervention to help teachers develop their own practices, a central component of the Daphne Project was the development of a PLC consisting of the authors and seven biology teacher participants, with a biologist providing content-specific guidance when necessary. These multiple stakeholders brought myriad perspectives to the PLC, allowing for the active exchange of ideas based on the participants' diverse expertise. The PLC met monthly during the school year for the duration of years two and three of the project. According to DuFour (2004), three components of successful PLCs in schools are (1) a focus on student learning, (2) a culture of collaboration and (3) a focus on results. Each of these components was undertaken within the Daphne Project work as the activities of the PLC were modeled on problem-solving and assessment development cycles from mathematics education (Borko, 2004; Webb, Romberg, Burrill, & Ford, 2005) to facilitate teachers' reflection on their current practices and adoption of new practices supported by the educative LP.

The Daphne Project PLC is modeled on an iterative four-step process. At each step in this process student learning was central in teachers' collaboration around specific instruction and feedback necessary to move students with specific prior ideas forward to correct understandings. The first step is for teachers to **Reflect** upon their current practice, sharing instructional strategies and materials with each other and discussing their advantages and disadvantages (Loughran, Mulhall, & Berry, 2004). In the second step, teachers **Explore** student thinking in their classes and compare those ideas with the LP. In this step, a particular focus is placed upon deepening teachers' content knowledge, as well as understanding students' common misunderstandings and alternative conceptions and the sources of those ideas. In the third step, teachers **Practice** using the associated tools and formative assessments, categorizing samples of student work, anticipating feedback, and discussing video cases. At this stage, encouraging teachers to discuss not only the identification of student ideas is essential, but also anticipating the kinds of feedback students within each category of student thinking would need. The fourth step has the teachers **Enact** their natural selection units with the tools contained in the educative LP, after which the teachers continue the cycle by reflecting upon their enactment and **Revising** their plans for next time.

## Method

### Participants

Eight biology teachers at Springfield High School participated. Springfield, considered among the top high schools in the state, is an ethnically and socioeconomically diverse school near a large city in the western US. The teachers represented a wide range of backgrounds and experience, ranging from a second-career student teacher to a 29-year veteran Biology teacher. Students in the participating teachers' classes were enrolled in four levels of biology: 9th grade pre-International Baccalaureate Biology, 10th grade General Biology, 10th grade English Language Learner (ELL) Biology, and 11th grade Advanced Placement Biology.

## Learning Progression and Professional Learning Community

On the basis of the student ideas about natural selection identified above, a draft LP was developed, piloted, and revised based upon analyses of student work (Furtak, 2009). By organizing students' ideas in a linear fashion, progressing from simple to complex, moving away from the ideas of Lamarck toward those of Darwin. As described above, the LP was accompanied by formative assessments designed to elicit student thinking, along with sample student responses and feedback strategies that were updated and extended by the reseachers and teachers through the course of the study.

Prior to the beginning of the study, the biology teachers at Springfield High did not have common planning time, and other than using the same textbook at each level of biology instruction, did not have a common curriculum. The teachers and the lead researcher participated in meetings every two to three weeks during the 2008-2009. Sessions involved discussing approaches to teaching evolution, predicting and exploring students' ideas about natural selection, categorizing student work, and discussing different instructional approaches given student ideas represented in the LPs.

## Sources of Data

Videotapes made of PLC meetings spanning the 2008-2009 and 2009-2010 academic years are the primary source of data for this study. Complimenting these videotapes are field notes kept by the authors of the study during each meeting, reflective notes made immediately followign each meeting, artifacts from the meetings, and content logs made of each videotape.

## Data Analysis

Through the course of the study, the authors kept ongoing notes based upond emergent themes they noticed during the PLC meetings. Reflection and discussion of these notes within research group meetings revealed converging themes centered on three shifts that occurred across the course of the academic year: shifts in the community interactions, shifts in the language teachers used to characterize student thinking, and shifts in practice.

The authors independently viewed videos supplemented by the artifacts described above and noted instances of interest in the PLC videos related to these three themes. Observations were categorized according to their relationship to the three elements of focus in this study: community, language, and instrucitonal practice. Then propositions were made on the basis of what each researcher viewed as the nature of the interactions in that video. The three researchers then met and came to a consensus about the nature of the three categories in each video. Longitudinal comparisons were then made across the set of PLC videos.

Table 2: Sample categories

| Categories | Questions to frame analysis |
|---|---|
| Community | How do teachers interact within the PLC community? Who allies with whom? What are the issues of status that arise? Who is speaking, and who is listening? How do they listen to and respond to each other? |
| Language | How do teachers characterize student ideas? What kinds of ideas do they mention, and what words are they using to describe them? How do these relate/not to the learning progression? Issues of creationism and intelligent design versus misconceptions? |
| Instructional Practices | What kinds of instructional practices are discussed? What are the rationales for using these teaching practices? How are learning experiences sequenced? How do these teaching practices relate to those represented in the learning progression? |

# Results

Results of our analyses of the multiple sources of data collected around the PLC meetings indicate that the PLC became more empathetic and communicative, respecting and exchanging ideas in a way that did not happen at the beginning of the study. In addition, teachers moved away from describing student beliefs in general terms and migrated toward description of the student ideas represented in the learning progression, using the common language present in the LP to describe and label these ideas. Finally, teachers' discussion of their instructional practices moved away from simple discussion of what their favorite activites were to more targeted description of which practices were better suited to the ideas represented in the LP. In the full paper, we will describe and illustrate how these changes occurred over the course of the study, describing in detail important events in this transformation process; however, for the purposes of the proposal, we summarize our results briefly below.

### Shifts in Community Dynamics

Early in the study, teachers took oppositional stances about practices in individual classrooms, without strong sense of a need for collaboration. For example, In the first PLC meeting, Lisa stated that that she was not trying to change student beliefs about evolution while Rachel stated that she didn't feel like she needed to respect non-scientific student ideas. Over time, however, the conversation in the PLC shifted almost completely away from oppositional stances and toward teachers expressing mutual respect for each other's arpproaches to teaching the strategies common to the LP, sharing strategies and their effectiveness at addressing particular student ideas, and taking up each other's ideas. For example, Rachel moved away from her oppositional stance later in the study, responding to a video of a colleague teaching with the statement, "I just really like that last bit of questioning she did…who would be the only predator…what would happen? So she, like, took it to be predictive and extended the idea." This comment highlights the development of mutual respect within the group. In addition, the balance of contribution to the PLC shifted away from its initial status, in which Rachel and Chris, the teachers most confident in their content knowledge, dominated the discussion at the expense of other teachers who were less confident, toward a more even balance of contribution. We argue that this shift took place not only because the previously dominant teachers gained a new respect for the practices of their colleagues, but also because the previously non-dominant teachers gained confidence in content knowledge through the PLC and therefore had more to contribute.

### Shifts in Language to Describe Student Ideas

Early in the study, teachers were not able to differentiate between different kinds of student ideas about natural selection. In early meetings, they spoke extensively either about students' rejection of evolution and natural selection or, when pushed to speak about common student understandings, teachers stated their impression of student ideas in general terms. This general stance is reflected in a statement by Alison, who said, "I just feel like yeah they're confused about all parts of natural selection so I'll just go from there." Later in the study, conversation about evolution all but disappeared in favor of explicit description of student ideas using the categories from the LP; for example, distinguishing between various categories of what we came to call 'need-based change' in the forms of students arguing that the environment induces changes in organisms, or that organisms mutate to fit their surroundings. In addition, teachers cited Mayr's facts and inferences when talking about the 'big ideas' students were expected to come to know. In fact, Chris and Lisa - teachers who previously had not worked collaboratively - met independently of the scheduled PLC meetings and categorized preassessments according to the categories in the LP.

### Shifts in Descriptions of Instructional Practices

We found that early in the study, teachers described which activities they liked to use to teach particular concepts, but did not have common activities that would facilitate their sharing of strategies across the group. Much of early discussion in the PLC was therefore focused on the teachers coming to an understanding of exactly what they were doing in their classrooms. However, as the study progressed, teachers agreed to use a common set of formative assessments in all of their classrooms as a basis for a discussion about their relative effectiveness in eliciting student ideas along the LP. In addition, teachers' conversations about the source of common student misunderstandings served as an impetus to create new formative assessments to which each teacher contributed, and which each teacher then used in her or her own classroom. In addition to these common activities that were a part of the educative LP, teachers came to discuss more explicitly which of the student ideas represented on the LP necessitated use of different activities in the classroom. Early in the development of the PLC teachers expressed hesitation to solicit student ideas about evolution. This hesitation appeared to be driven by a fear of creating conflict in the classroom. Later, this hesitation disappeared and teachers not only were interested in soliciting student ideas but also the sources of those ideas. We suggest that the confidence to tackle student ideas grew out of the stronger content knowledge and instructional strategies developed in the PLC.

In addition to these common formative assessments, teachers also came to share the different variations of common class activities that they were using in their natural selection units. For example, several teachers used some version of the Biological Sciences Curriculum Study's (2006) activity in which students pick up colored dots from patterend cloth, discovering over multiple iteractions of the activity that the harder-to-see dots were less likely to be picked up from the cloth. During the course of the study, all teachers decided to use some form of this activity, some using physical materials and others using an online applet to model the same phenomenon. Teachers then came together later in the study and discussed the relative effectiveness of the activity, weighing trade-offs such as time spent on the activity and depth of student understanding.

Another change in the discussion of instructional practices within the PLC was a change in focus from talking about activities to specifically identifying and addressing student misconceptions, and targeting which instructional examples, analogies, and activities would best address them. In addition, teachers reflected on the fact that students reverted back to their naïve ideas when presented with novel scenarios in class. This highlights the fact that these teachers, in addition to developing more advance strategies for eliciting and identifying student ideas, are also reflecting upon the process of conceptual change in a more sophisticated manner than at the beginning of the study (Smith, diSessa, & Roschelle, 1993).

.

## Educational Importance

Thte preceding data indicate that the PLC centered on the educative LP provided teachers an opportunity to reconnect and develop a common language around a single conceptual domain, enriching their own content knowledge while they simultaneously built a new community in their department. We found that repeated meetings centered on the educative LP helped to refocus teachers from defensive stances regarding the teaching of evolution to the smaller instrucitonal issue of natural selection, to enrich teachers' common language regarding student ideas, and inform their discussion and selection of instructional strategies targeted at student thinking.

The present research also illustrates the importance of involving the 'user' audience in the development of LPs and their accompanying tools.

Given the current level of interest in the development and study of LPs in science education, the present analysis is of critical importance to inform educational researchers and professional developers of the utility of these tools not only in guiding assessment and curriculum development, but also in developing the content knowledge of practicing science teachers and building community around a common content area.

## References


American Association for the Advancement of Science. (2001). *Atlas of Science LIteracy*. Washington, D.C.: National Academies Press.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and Evaluation of the Conceptual Inventory of Natural Selection. *Journal of Research in Science Teaching, 39*(10), 952-978.

Biological Sciences Curriculum Study. (2006). *BSCS Biology: An Ecological Approach* (10th ed.). Dubuque, Iowa: Kendall/Hunt.

Bishop, B. A., & Anderson, C. W. (1990). Student Conceptions of Natural Selection and Its Role in Evolution. *Journal of Research in Science Teaching, 27*(5), 415-427.

Black, P. (1998). Assessment by Teachers and the Improvement of Students' Learning. In B. J. Fraser & K. G. Tobin (Eds.), *International Handbook of Science Education* (pp. 811-822). Dordrecht, Netherlands: Kluwer Academic Publishers.

Brown, A. 1992. Design Experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of the Learning Sciences, v.2, n.2, p.141-178.

Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. *Educational Researcher, 33*(8), 3-15.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic Assessment with Multiple-Choice Items. *Educational Assessment, 11*(1), 33-63.

Catley, K., Lehrer, R., & Reiser, B. (2005). *Tracing a Prospective Learning Progression for Developing Understanding of Evolution*: Paper Commissioned by the National Academies Committee on Test Design for K-12 Science Achievemento. Document Number)

Collins, A. (1998). National Science Education Standards: A political document. *Journal of Research in Science Teaching, 35*(7), 711-727.

Davis, E. A., & Krajcik, J. (2005). Designing Educative Curriculum Materials to Support Teacher Learning. *Educational Researcher, 34*(3), 4-14.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher, 35*, 125-129.

Driver, R., Guesne, E., & Tiberghien, A. (1985). *Children's Ideas in Science*. Philadelphia: Open University Press.

DuFour, R. (2004). What is a "Professional Learning Community"? *Educational Leadership*, 61(8), p. 6-11.

Duncan, R. G., Rogat, A., & Yarden, A. (2009). A learning progression for deepening students' understanding of modern genetics across the 5th-12th grades. . *Journal of Research in Science Teaching, 46*(6), 655-674.

Duschl, R. A. (2003). Assessment of Inquiry. In J. M. Atkin & J. Coffey (Eds.), *Everyday Assessment in the Science Classroom* (pp. 41-59). Arlington, VA: NSTA Press.

Ferrari, M., & Chi, M. T. H. (1998). The nature of naive explanations of natural selection. *International Journal of Science Education, 20*(10), 1231-1256.

Furtak, E. M. (2009). *Toward Learning Progressions as Teacher Development Tools*. Paper presented at the Learning Progressions in Science Conference, Iowa City, IA.

Gess-Newsome, J. (1999). Secondary Science Teachers' Knowledge and Beliefs about Subject Matter and Their Impact on Instruction. In J. Gess-Newsome & N. G. Lederman (Eds.), *Pedagogical Content Knowledge and Science Education* (pp. 51-94). The Netherlands: Kluwer Academic Publishers.

Lieberman, A, and Miller, L. (2008). "Balancing content and process: Facing challenges." In Lieberman, A. and Miller, L. (eds.), *Teachers in Professional Communities*. New York: Teachers College Press, p.29-38.

Loughran, J., Mulhall, P., & Berry, A. (2004). In Search of Pedagogical Content Knowledge in Science: Developing Ways of Articulating and Documenting Professional Practice. *Journal of Research in Science Teaching, 41*(4), 370-391.

Luft, J. A., & Roehrig, G. H. (2007). Capturing Science Teachers' Epistemological Beliefs: The Development of the Teacher Beliefs Interview. *Electronic Journal of Science Education, 11*(2), 38-63.

Mayr, E. (1997). *This is Biology*. Cambridge, MA: Harvard University Press.

McLaughlin, M. & Talbert, J. (2006). Building school-based teacher learning communities: professional strategies to improve student achievement. New York: Teachers College Press.

National Research Council. (1996). *National Science Education Standards*. Washington, D.C.: National Academy Press.

National Research Council. (2001). *Classroom Assessment and the National Science Education Standards*. Washington, D.C.: National Academy Press.

National Research Council. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8.* Washington, D.C.: National Academies Press.

Rudolph, J. L., & Stewart, J. (1998). Evolution and the Nature of Science: On the historical discord and its implications for science education. *Journal of Research in Science Teaching, 35*(10), 1069-1089.

Rutledge, M. L., & Mitchell, M. A. (2002). High School Biology Teachers' Knowledge Structure, Acceptance, and Teaching of Evolution. *The American Biology Teacher, 64*(1), 21-28.

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science, 18*, 119-144.

Sandoval, W. & Bell, P. (2004). Design-based research methods for studying learning context: introduction. Educational Psychologist, v.39, n.4, p.199-201.

Shavelson, R. J. (2009). *Reflections on Learning Progressions*. Paper presented at the Learning Progressions in Science Conference.

Shepard, L. (2009). *From Advocacy to Adoption: Getting Formative Assessment into More Classrooms. Discussant Remarks*. Paper presented at the American Educational Research Association Annual Meeting.

Shtulman, A. (2006). Qualitative differences between naive and scientific theories of evolution. *Cognitive Psychology, 52*, 170-194.

Shulman, L. S. (1986). Those Who Understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Sloane, F. (2009). Randomized trials in mathematics education: recalibrating the proposed high watermark. *Educational Researcher,* 37(9), p.624-630.

Smith, J. P., DiSessa, A. A., & Roschelle, J. (1993). Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Webb, D. C., Romberg, T. A., Burrill, J., & Ford, M. J. (2005). Teacher Collaboration: Focusing on Problems of Practice. In T. A. Romberg & T. P. Carpenter (Eds.), *Understanding Math and Science Matters* (pp. 231-251). Mahwah. NJ: Lawrence Erlbaum Associates.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: University of Cambridge Press.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, N.J.: Erlbaum.

Wilson, M. (2009). *Learning Progressions to Guide Systems of Formative and Summative Assessment*. Paper presented at the Annual Meeting of the American Educational Research Association.

Windschitl, M. (2004). Folk theories of "inquiry:" How preservice teachers reproduce the discourse and practices of an atheoretical scientific method. *Journal of Research in Science Teaching, 41*(5), 481-512.

# Measuring Transformative Modeling: A Framework of Formatively Assessing Students' Deep Conceptual Understanding in Physical Sciences

Ji Shen, University of Georgia, Mathematics and Science Education, Athens, GA, jishen@uga.edu
Ou Lydia Liu, Educational Testing Service, Princeton, lliu@ets.org
Hsin-Yi Chang, National Kaohsiung Normal University, Graduate Institute of Science Education, Kaohsiung, Taiwan, hsinyichang@nknucc.nknu.edu.tw

**Abstract:** Measuring students' conceptual understanding can provide important information for learning and teaching. Many formative assessment approaches elicit students' prior knowledge without identifying the sources of their misconceptions. This proposal presents a framework that guides the development of formative assessment to measure student understanding of physical science topics using three types of connections. Particularly, we assess how students link physical states, processes and mechanism, integrate different scientific models, and connect scientific knowledge to everyday experience. Based on this framework, we construct sample items on sinking and floating and piloted the items with 18 preservice middle grades science teachers. Analyses of student responses to these sample items provide evidence of the effectiveness of the framework in extracting information valuable to improve learning and instruction. Criteria and implications of using this formative assessment framework in science classrooms are then discussed.

## Introduction

Many national and state-level science tests fail to measure deep understanding (Hyde et al. 2008; Liu, Lee, Hoftstetter, & Linn, 2008). As a result, American students continue to languish on international science assessments (OECD, 2006). To reverse this trend and promote science achievement in the U.S., it is critical to design formative assessments that can effectively capture students' prior experience and measure deep understanding (Shen, Gibbons, Wiegers, & McMahon, 2007).

The challenge of capturing students' deep understanding through formative assessment is manifold. The complicated dimensions of science learning create challenges in both assessment design and utility of the information extracted from the assessment. For instance, many science assessments targeting conceptual understanding are designed to measure a single aspect of the correctness of science content. However, science learning is such an integrated and dynamic system that multidimensional abilities should be considered in the assessment (Anderson & Krathwohl, 2001). The assessment should capture student ability in integrating understanding and justifying explanations as well as their mastery of content knowledge. Most importantly, to take advantage of the assessment results, teachers should understand how students apply textbook knowledge in a new learning context.

To address these challenges, many science education researchers have developed frameworks to capture the multiple dimensions of science learning that shifted away from only focusing on content correctness (Linn et al. 2006; Yin et al. 2005). Science educators are also interested in understanding how students develop competences such as inquiry and modeling abilities to conduct science experiments (e.g., Gotwals & Songer, 2006). Acknowledging the multiple dimensions of science learning, we expand the dimension of conceptual understanding to capture students' modeling and inquiry abilities in this study. Particularly, we present a framework that guides the development of formative assessment to measure students' understanding of abstract and complex topics in physical sciences in three types of connections: linking physical processes and states with underlying mechanism, integrating multiple explanatory models, and connecting their science knowledge to natural observations and everyday experiences (National Research Council [NRC], 2000; Shen & Linn, 2010).

We focus on formative assessments that help teachers identify students' weaknesses in order to learn science concepts more effectively. Our framework can be enacted in multiple formats, e.g., paper and pencil quizzes, computerized word problems, oral interviews, or group-discussion questions. In the following section, we first discuss some exemplar assessment approaches in science education on which our framework is built. We then present the rationale and mechanism of our framework. Finally we elaborate on the framework with sample student responses on an item in sinking and floating.

## Measuring Deep Conceptual Understanding

Effective assessment practices in science education are aligned with distinctive learning theories (NRC, 2001; Wilson, 2005). Here we elaborate on three examples that influenced the development of our framework.

### Measuring Integrated Understanding Using Knowledge Integration

Many students gain fragmented knowledge. Linn and colleagues have developed a knowledge integration (KI) framework that emphasizes student abilities in establishing connections among ideas (Linn & Eylon, 2006; Linn et. al., 2006; Liu et al. 2008). The KI framework emphasizes the repertoire of ideas that students build as they interact with the world. Students add varied ideas as they experience science in daily life. The KI framework calls for taking advantage of the reasoning that students employ to formulate these views. The framework promotes coherent understanding by encouraging students to add new ideas, distinguish new and existing ideas, develop scientific criteria to reconcile ideas, and build coherent connections. The KI framework is realized through technology-enhanced learning environments that present complex and usually unseen science topics through computer visualization programs. The KI assessment framework features multiple assessment formats (e.g., pre/post tests, annual assessment, embedded assessment) for both formative and summative evaluation purposes. The KI scoring rubric rewards student reasoning and ability to articulate scientific evidence as well as content mastery. The KI assessment framework has been tested with large-scale empirical studies and has demonstrated satisfactory psychometric properties (Lee, Liu, & Linn, in press; Liu et al. 2008).

The KI framework states that there are different types of connections (Linn & Eylon, 2006) but does not distinguish these types in scoring (Linn et al. 2006). Building on this framework, we further detail and speculate that three types of connections are extremely important (will be discussed momentarily) in helping students gain deep conceptual understanding across science disciplines. Our assessment framework is aimed at identifying these different types of connections, which in turn can inform instructions.

## Measuring Understanding of Scientific Models

Modeling-based instruction in science classrooms is a constructivist approach that encourages students to develop, use, and evaluate models to describe, explain, and predict scientific phenomena (Chinn & Samarapungavan, 2008; Clement, 2000; Lehrer & Schauble, 2006; Schwarz & White, 2005). For example, through the MoDeLS project, Schwarz et al. (2009) developed a learning progression framework based on modeling for upper elementary and middle grade students. They defined a scientific model as "as a representation that abstracts and simplifies a system by focusing on key features to explain and predict scientific phenomena" (Schwarz et al. 2009, p.633). They considered two dimensions of the practice of modeling. One dimension addressed scientific models as tools for prediction and explanation; the other dimension considered the change of models when understanding improves. They built a construct for assessing students' progress in each dimension.

The works by Schwarz et al. (2009) and others provide important insights in how to teach and measure modeling in science classrooms. Our work focuses primarily on how deeply students understand scientific models and how well students are able to connect scientific models to natural phenomena. We believe that robust understanding of existing models can help students compare, evaluate, and revise models. We agree with Schwarz et al. (2009, p.635) in that "it is crucial to involve learners in the construction of models." However, we argue that it is equally important that students firmly grasp provided models because these models provide the solid foundations from which students can construct their own ones (Shen & Confrey, 2007). When students learn more advanced models in higher grades, they need to build on and integrate the simplified ones they have learned earlier on. Therefore, one important component in our framework is to assess how students integrate different scientific models they have learned.

## Measuring Complex Reasoning in Scientific Inquiry

Science education researchers also look at students' complex reasoning in inquiry. Principled Assessment Designs for Inquiry (PADI) is an assessment framework that focuses on providing design patterns and structures to measure students' inquiry skills and science knowledge. For example, Gotwals and Songer (2006) created a content-inquiry matrix that laid out different levels of science inquiry and science content required in assessment tasks. They used the matrix to highlight the interplay between inquiry abilities and science knowledge as students performed inquiry tasks or took inquiry assessments. The matrix had three levels of science content knowledge and three steps of science inquiry processes. The simple level of science content knowledge meant that the task or assessment provided sufficient content; the moderate level required students to demonstrate solid understanding of science concepts; the complex level required students to connect among multiple concepts. For the three steps of science inquiry the authors indicated that the steps might differ depending on the aspects of inquiry being targeted, such as generating scientific explanations or constructing representations using data.

While the content-inquiry matrix is used to help assess how students develop science knowledge and inquiry skills at different levels of complexity over time, our framework focuses on laying out important aspects of conceptual understanding that students need to explain complex phenomena specifically in physical sciences. Similar to the PADI framework, our framework assesses complex reasoning closely related to inquiry such as explaining phenomena in light of its underlying mechanism, or connecting between multiple explanatory models. The difference is that our framework can be used in a variety of teaching and learning contexts that may or may not necessarily measure the full scale of inquiry.

We propose a new assessment framework on conceptual learning to address the limitations that we see in the existing approaches. Built upon the existing literature, our work is an ongoing contribution to advance the formative assessment of students' deep understanding in science classrooms. We envision such a framework as a useful tool for instructors to revamp instruction based on the information extracted from the assessment responses.

## Theoretical Framework

Our framework is informed by the instructional theory of transformative modeling (TM) (Shen & Confrey, 2007, 2010). It is a theoretical framework used to describe, analyze, and inform learning processes. TM delineates learning and teaching as a process of modeling the natural world through chains of operations on materials. At the center of the operations are a set of transformations that alter the nature of physical or symbolic objects by adding or suppressing information. The transformed materials, as well as the operations on these materials, render potential for future learning. Examples include transforming detailed observations into numerical records to ground scientific understanding of measurement uncertainty, transforming data and graphs into mathematical formula to enrich abstract conception of motion, transforming between the geocentric model and the heliocentric model to better comprehend the solar system.

In this framework, conceptual learning is defined as the process of gaining ability to transform materials in both physical and abstract forms. Specifically, the consistency among the transformed materials perceived by a cognizing agent defines the depth of the comprehension of the concept. The variety of transformation types and forms determines the broadness of the construed concept. For the depth of comprehension, an example is the common misconception that many students hold about the cause of the seasons (Harvard-Smithsonian Centre for Astrophysics & Schneps, 1988). Many students reason that it is hot in the summer because the sun is closer and it is cold in the winter because the sun is farther away (the distance model). This reasoning draws upon everyday experience (e.g., sitting besides a fireplace) that one feels warmer when sitting closer to a heat source. In the TM theory, we interpret this explanation as being able to transform everyday experience to explaining the seasons to a certain degree, even though it is not compatible with the scientific model. People later learn that the northern hemisphere and the southern hemisphere have opposite seasons, and that in fact in summer the earth is farther away from the sun. The distance explanatory model then becomes conflicting to the new information, which calls for a deeper understanding of the cause of the seasons. For the broadness of conceptual understanding, the TM framework advocates the incorporation of multiple representations in learning science to ensure coherent understanding. For instance, when students learn kinematics, constant transformation among graphic representations can help them summarize many observations and physics principles (Shen, 2009). Students need to acquire a rich set of graphs (e.g., position-time graph, velocity-time graph, free-body diagram) to broaden their understanding of Newtonian mechanics.

Based on the TM instructional theory and previous analyses of students' responses to questions on electrostatics (Shen & Linn, 2010), we identify three most important aspects in knowledge transformations: (a) linking scientific mechanism with both states and processes of the target domain; (b) integrating multiple explanatory models, and (c) connecting scientific models with everyday observations. In the following we elaborate on these three important transformations.

### Linking Physical States, Processes, and Underlying Mechanism

The idea of integrating mechanism with states and processes comes from the fact that much scientific understanding originates from transforming observations of natural changes into data, inscriptions, and digital records. During such an observation, the states and the processes need to be accurately described. The scientific records of the initial, transitional, and final states help students identify the interacting agents and the relevant variables that constitute the key properties of the target system. A physical process of a system consists of changing states over time. Many of these processes lead to a physical equilibrium (e.g., thermal equilibrium) in which the state of the system does not change anymore. A mechanism is defined as an explanation (e.g, mathematical formalism, causal inference) one draws to theorize the patterns of how the key states emerge and change. Oftentimes, the mechanism introduces abstract concepts (e.g., force) or microscopic entities (e.g., electron).

Here are two examples in physical sciences that illustrate how students need to integrate mechanism with states and processes to gain deep understanding. An electrical *state* of an object may be positively charged, negatively charged, or neutral. For example, an object is initially positively charged. Then, it is discharged after touching one's hand, and eventually, the object becomes neutral. Students may only pay attention to the initial state (positively charged) and the final state (neutral) of the object without thinking about the in-between *process* when the change takes place. The understanding of the charging/discharging processes provides a trajectory of the observed changes and trends of the electrical states. Therefore, a better understanding includes the process where the negative charges are transferred from the hand to the object to neutralize it. The understanding is further enhanced by learning the *mechanism* of why charges move: like charges repel and opposite charges attract. That is, the forces between charges drive the motion of the charges. Likewise, deep understanding of the state of thermal equilibrium involves explaining the heat transfer processes. A thermal state of an object (hot, warm, or cold) can be measured by its temperature. Consider a hot object in touch with a cold object. The initial thermal states of the two objects are measured in different temperatures. Many students learn that, eventually, the two objects will reach the same temperature. They need to know how this happens by learning the processes and mechanism. When envisioning that heat is transferred from the hot object to the cold object, one is considering the process. A mechanism states that temperature difference drives heat to transfer from an object with higher temperature to another object in touch until they reach the same temperature (thermal equilibrium).

### Integrating Multiple Explanatory Models

One may build different models to explain the same phenomenon since there are different aspects of the phenomenon one chooses to model (Shen & Confrey, 2010). Some of the models are consistent with the accepted scientific models; others conflict with accepted understanding. Students typically start with simple and concrete models, and then advance to more abstract and general ones in learning complex science topics (Lehrer & Schauble, 2006). We argue that when learning more advanced models, students need to integrate the earlier models they have learned. For instance, to explain and predict how the planets move in the solar system (kinematics only), students may start with a geocentric model since that is what one sees from the earth. Then students may move to learn the heliocentric model used to explain the geocentric model. A higher-level understanding indeed incorporates the two models and helps students go back and forth between the two models depending on the context.

For the discharging example, one may employ a charge-based model as discussed in the previous section. One may also use a particle-based model: initially, the object has less electrons (which carry negative charges) than a neutral state (the final state). During the discharging process, electrons from human hand move to the object. The movement of electrons is driven by the interactions (forces) among the particles. The connection between the two models lies in the fact that types of charges are properties of certain particles (e.g., electrons are negatively charged). For thermal equilibrium, the heat transfer process may be attributed to a substance carrying heat that flows from one object to another, which may be considered as a misconception. Alternatively, one may consider a particle model. In this model, temperature of an object is a measure of the average kinetic energy of the particles, which corresponds to the average speed of the particles. That is, an object with a higher temperature has particles vibrating faster on average than an object with a lower temperature. In the heat transfer process, collisions between particles drive the transfer of the kinetic energy from one object to the other.

### Connecting Science Knowledge to Physical Observations and Everyday Experience

Physical sciences study the patterns of the physical world and students experience the physical world every day. It is expected that students learn science more effectively and meaningfully if their learning is built upon and tied back to their everyday experience (Linn, 2006). Unfortunately, many teachers do not use students' experience as resources for teaching. As a result, students often build models in the learning context but rarely connect to their learning experiences in other contexts or to everyday experiences (Kozma, 2003). Being able to connect scientific models to other learning contexts or everyday experiences is an indicator of the degree to which students develop integrated understanding of the represented concept.

Many examples of everyday experience are related to discharging: e.g., grabbing a metal doorknob and getting shocked, touching the metal part of a car before refueling, spraying water to prevent fluffy hair. Deep understanding of electrostatics requires students to make sense of these phenomena outside of classrooms. Similarly, when learning heat and temperature, students should reflect on everyday experiences such as determining a better material to keep things warm.

## Constructing Assessment Items on Sinking and Floating

Sinking and floating are interesting phenomena for students to investigate. It involves different explanatory models and everyday experience. Abundant previous research was devoted to identify students' understanding of sinking and floating (e.g., Dentici et. al. 1984; Halford, Brown, & Thompson, 1986; Mullet & Montcouuiol, 1988; Rowell & Dawson, 1977; Yin, Tomita, & Shavelson, 2008). In the following section, we first introduce the physics behind the phenomena of sinking and floating. We then present three sample items to illustrate how we apply the TM framework to build assessments that help students identify the connections among key states and processes, link multiple explanatory models, and apply to the physical world. Finally we present sample student responses and demonstrate how the responses can be analyzed using our framework.

## Physics of Sinking and Floating

To determine if an object is either a sinker or a floater without putting it in a fluid, one can compare its average density with the density of the fluid - the density model. If the average density of the object is greater than the fluid, then the object will sink when put in the fluid. If it is smaller, the object will float. If the average density of the object is equal to the fluid, then the object will suspend in the fluid. Alternatively, one may also compare the weight of the object with its buoyancy - the force model. Buoyancy on an object is defined as the net force exerted by the surrounding fluid. [(1)] One may simply predict that an object will sink when put in the fluid if the buoyant force on the object is smaller than its weight. But the opposite is only partially true: the object will *start to* float if its buoyancy is greater than its weight. This only applies for an object completely submerged under water. It may lead to the misconception that the buoyancy is greater than weight when an object floats on water. Having learned the density model, one may think that the force model is redundant. In fact, the force model is able to explain the processes of sinking and floating in much more details. [(2)]

A deep understanding needs students to be able to integrate the density model and the force model. To compare the buoyancy and the weight of an object completely submerged under a fluid, one may compare its weight with the weight of the fluid displaced since the buoyancy of an object is equal to the weight of the fluid displaced (Archimedean principle). Also since the object is immersed in the fluid, the volume of the object is the same as the volume of the fluid displaced. Hence, it is equivalent to compare the density (mass/volume) of the object with the density of the fluid.

## Instruments Used to Assess Understanding of Sinking and Floating

Figure 1 shows a sample item in an instrument used to identify students' misconceptions developed by Yin, Tomita, and Shavelson (2008). This item does a nice job in revealing students' misconception, i.e., "Big/heavy things sink, small/light things float." Some students may predict the two blocks will "sink" when stacked together (or subsurface float) because the bundle of blocks A and B is heavier than each individual block.

Block A and Block B both float in water. Suppose that we glue them firmly together and place them in water; together they will: ____________.

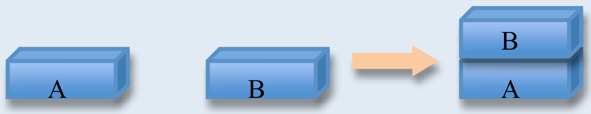



Figure 1. A sample item on sinking and floating from Yin et al. (2008).

However, this kind of items does not inform the teacher of the possible origins of these misconceptions. Using our TM framework, a possible explanation for the misconception is that the students may not distinguish density (mass over volume) from weight (a force). This explanation is also consistent with our experiences of teaching the topics on sinking and floating. In everyday language, the term heaviness often refers to the two terms interchangeably. [(3)] One can explain the phenomenon of sinking and floating using the density model or the force model (detailed explanation is provided in the next section). But these explanatory models may be taught in a fragmented fashion, and students are not able to connect these models. Moreover, this kind of items does not inform the teacher if students are able to connect the knowledge of sinking and floating to observations. We may infer that the incorrect answer entails that the student cannot connect to everyday experience, e.g., if one bundles two pieces of wood together, the bundle still floats.

## A Sample Item on Sinking and Floating Using the TM Framework

*Participants.* We created an instrument on sinking and floating using the TM framework and piloted with 18 undergraduate students working towards middle grades science teaching certificates (17 female, 1 male). These students had weak science background based on the results of a survey measuring general physics knowledge (25% correct). The items on sinking and floating were administered in a quiz after they had two lessons on sinking and floating (4 hours in a total) including a lab confirming Archimedean Principle. The quiz has six constructed-response items. It took the students half to one hour to finish. Here we provide an example.

*Item Two Balls* [**due to space limit, here we only present one sample item and students' responses**]
The item Two Balls (see Figure 2) asks students to predict and explain what will happen to two balls (one floater and one sinker) in water when salt is added. A physical demonstration can also be set up. Students can use either the density model or the force model to explain the observations as in the previous item. The key points are summarized in Table 1.

Ball A and ball B have the same volume. A is floating on water, B is sitting at the bottom of the water tank.
(1) Which ball has the greater buoyancy? Explain.
(2) Which ball has the greater density? Explain.
(3) Sam adds some amount of salt in water and let them dissolve. What happens to the buoyancy of the two balls? Explain.
(4) What changes would you expect after Sam has added salt in the water?
[A physical demonstration can be shown when students respond to this problem].

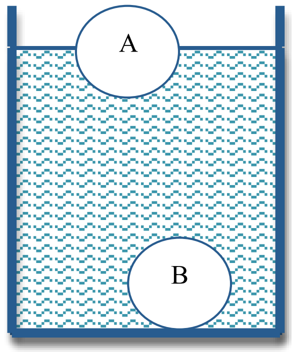



Figure 2. Item "Two Balls" on sinking and floating.

Table 1: The density and force models used to explain the item Two Balls.

| | Density Model | Force Model |
|---|---|---|
| State | • Ball A is floating initially. Part of ball A is submerged under water.<br>• Ball B is sitting at the bottom initially. All of ball B is submerged under water.<br>• After adding salt, ball A still floats. Ball B may sink or float depending on the amount of salt added in.<br>• After adding salt, the density of the water increases. | • Ball A is floating initially. Part of all A is submerged under water.<br>• Ball B is sitting at the bottom initially. All of ball B is submerged under water.<br>• After adding salt, ball A still floats. Ball B may sink or float depending on the amount of salt added in.<br>• After adding salt, the weight of the water of the same volume increases. |
| Process | • When the salt is added, more part of ball A comes out of water.<br>• When salt is added, ball B may still sit at the bottom. It may also move up and becomes a floater. | • When the salt is added, more part of the ball A comes out of water.<br>• When salt is added, ball B may still sit at the bottom. It may also move up and becomes a floater. |
| Mechanism | • An object less dense than water tends to float. An object denser than water tends to sink. An object having the same density as water suspends in water.<br>• For a floater, the proportion of the volume of the object that is submerged under water over its whole volume equals the proportion of the density of the object over the density of the fluid (relative density). | • An object accelerates if the net force acting on it is non-zero (Newton's second law). An object is at rest or moves at a constant velocity if the net force acting on it is zero (Newton's first law).<br>• There are two forces acting on an object floating on water: weight and buoyancy.<br>• There are three forces acting on an object sitting at the bottom of a water tank: weight, buoyancy, and normal force by the tank.<br>• Buoyancy is equal to the weight of fluid displaced (Archimedean Principle). |

The detailed explanation follows. Since ball A displaces less volume of water than ball B, it has a smaller buoyancy than B. For a floater, its density is less than the density of water; for a sinker, its density is greater than the density of water. So the density of ball A is smaller than that of ball B. When salt is added, the density of the salt water is increased. Therefore, ball A keeps floating. For any floater, the buoyancy equals its weight, so the weight of ball A does not change. This means that the buoyancy of ball A does not change. To maintain constant buoyancy, ball A has to rise up a bit to displace less volume of salt water. For ball B, it may still rest at the bottom, but exerting less force on the bottom of the water tank. It may also move up if the amount of salt dissolved leads to a greater density of the salt water than ball B. If that is the case, Ball B will go all the way to the top until a part of it comes out of the salt water.

*Students' Responses and Implications for Instruction*

All students stated that ball B has a greater density in sub-question #2. This indicates that all the students are able to use the density model to explain the final states of sinking and floating at this stage. However, students' responses to sub-questions #1, #3 and #4 revealed that most of them did not integrate the buoyancy model to explain and predict the phenomenon of sinking and floating.

Students may confuse the sinking and floating states with the underlying mechanism of competing forces. Buoyancy is a variable (force) that influences the sinking or floating of an object. However, students often think that buoyancy refers to the state of floating (i.e., floating higher indicates a greater buoyancy), and may think that buoyancy is a property of an object. In the pilot test, 10 out of 18 students thought that ball A had a greater buoyancy in sub-question #1. A typical response stated that "Ball A has the greater buoyancy because it is floating on the water." Another student wrote, "Ball A has a greater buoyancy because it has a density that is less than water and is floating, (therefore) buoyant forces acting on it are greater." It was very likely that this student had a concept of "relative buoyancy" in mind. This student might equate buoyancy to the relative magnitude of buoyancy and its weight, corresponding to the relative density of the object compared with the density of the water. Students often ignore associated changes when one change is introduced to a physical system (Shen et al. 2007). For sub-question #3, all students responded that both balls changed in the same manner: the two balls either had a greater buoyancy, smaller buoyancy, or the same buoyancy. Similarly, for sub-question #4, 12 out of 18 students responded that both balls floated "better" or "higher." All of them recognized the change that when salt was added, the density of the fluid was increased. However, most of them didn't distinguish the two balls as two different cases, as one student reasoned, "the two balls will float better in the salt water, since the difference between the densities of the balls and salt water is greater, both balls will float higher than they did in plain water."

Analyses of these responses provide many insights on how to help students gain deep understanding of sinking and floating in future instruction. These items create the need and the instructional means for students to integrate different models they have learned, as they may realize through class discussion or instructors' feedback that their explanatory models conflict with each other. Follow-up questions that target detailed changing processes of sinking and floating may help students to connect these models. Sample questions may include "how does ball A 'know' how much to rise up when salt is added? How does ball B 'know' when to go up? What happens to its buoyancy when ball B moves up? When ball B comes out of water, does it stop right away?" These questions also call for more careful observations that help the students link the states and processes of sinking and floating. Students' alternative conceptions and prior experience can serve as knowledge resources for bridging different explanatory models. For instance, conceivably, students' understanding of "relative buoyancy" as the ratio of buoyancy over its weight may lead to the scientific understanding: for an object less dense than water, it has a greater ratio of buoyancy over weight.

## Conclusion and Implication

In this study we present an assessment framework that focuses on three aspects of conceptual understanding: linking physical states and processes with underlying scientific mechanism, integrating multiple explanatory models, and connecting scientific models with everyday experience or natural observations. We illustrate with students' responses to items on sinking and floating. Careful analyses show that these aspects are imperative for students to learn science. Specifically, the proposed framework has potential to help instructors identify the important links that students often miss among multiple explanatory models. With information provided by the framework instructors can effectively observe the detailed processes that students need to explain certain science phenomena, and help students integrate prior knowledge or experience to new science concepts.

## Endnotes

(1) From a microscopic view, the net effect of buoyancy is due to the collisions of numerous fluid particles -- the particle model, which is not discussed in this paper.

(2) Considering an object completely submerged under a fluid. If its buoyancy is greater than its weight, then the object starts to move up in acceleration until it floats above the fluid. When the object moves out of water, it displaces less water, hence experiences less buoyancy. When the object displaces the right amount of water so that its buoyancy equals its weight, the object does not stop immediately. It has acquired momentum, so it keeps moving out of water. Then the weight takes over and the object is pulled back towards the water. After some oscillations at the surface of the water, it stops eventually and its buoyancy and weight are balanced. This force model connects to Newtonian mechanics (force corresponding to acceleration) and links process (moving up and down) and with initial and final states (floating or sinking). The density model also connects to the mathematical construct of ratio, which has many implications.

(3) Most of them are also not able to distinguish mass and weight, which is relevant, but probably not critical here.

## References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Chinn A.C., & Samarapungavan A. (2008). Learning to use scientific models: Multiple dimensions of conceptual change. In: R.A. Duschl & R.E. Grandy (Eds.), *Teaching Scientific Inquiry* (pp 191-225). Rotterdam, Netherlands: SensePublishers.

Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, *22* (9), 1041-1053.

Dentici, O. A., Grossi, M. G., Borghi, L., Ambrosis, A. D., & Massara, C. I. (1984). Understanding floating: A study of children aged six and eight years. *European Journal of Science Education, 6*(3), 235-243.

Gotwals, A., & Songer, N. (2006). *Cognitive Predictions: BioKIDS Implementation of the PADI Assessment System (PADI Technical Report 10).* Menlo Park, CA: SRI International.

Halford, G.S., Brown, C.A., & Thompson, R.M. (1986). Children's concepts of volume and flotation. *Developmental Psychology, 22,* 218–222.

Harvard-Smithsonian Centre for Astrophysics (Prod.), & Schneps, M.H. (Dir.). (1988). *A private universe.* Santa Monica, CA: Pyramid Film & Video.

Hyde, J.S., Lindberg, S.M., Linn, M.C., Ellis, A.B., Williams, C.C. (2008). Gender similarities characterize math performance. *Science*, *321*, 494-495.

Kozma, R. B. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction, 13*, 205-226.

Lee, H. -S., Liu, O. L., & Linn, M. C. (in press). Validating Measurement of Knowledge Integration Science Using Multiple-Choice and Explanation Items. *Applied Measurement in Education*.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R.K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371-388). New York: Cambridge University Press.

Linn, M.C. and Eylon, B.-S. (2006). Science education: Integrating views of learning and instruction. In P.A. Alexander and P.H. Winne (eds.) *Handbook of Educational Psychology (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M.C., Lee, H.-S., Tinker, R., Husic, F., & Chiu, J.L. (2006). Teaching and assessing knowledge integration in science. *Science*, 313, 1049-1050.

Liu, O. L., Lee, H.S., Hoftstetter, C. & Linn, M.C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment*, *13*, 33-55.

Mullet, E., & Montcouquiol, A. (1988). Archimedes' effect, information integration and individual differences. *International Journal of Science Education, 10 (3),* 285–301.

National Research Council. (2000) *Inquiry and the national science education standards: A guide for teaching and learning.* Washington, DC: National Academy Press.

National Research Council. (2001). *Knowing what students know: the science and design* of *educational assessment*. Washington, DC: National Academies Press

Organization for Economic Co-operation and development (OECD) (2004). *Problem solving for tomorrow's world: First measures of cross-curricular competencies from PISA 2003*. Paris: Author.

Rowell, J.A., & Dawson, C.J. (1977). Teaching about floating and sinking: An attempt to link cognitive psychology with classroom practice. *Science Education, 61,* 245–253.

Schwarz, C.V., & White, B.Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, 23(2), 165-205.

Schwarz, C.V., Reiser, B.J., Davis, E.A., Kenyon, L., Acher, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners, *Journal of Research in Science Teaching*, 46(6), 632-654.

Shen, J. (2009). Walking out graphs: Multiple representation isn't enough. *Science Scope*, 33 (4), 47-51.

Shen, J. & Confrey, J. (2007). From conceptual change to transformative modeling: A case study of an elementary teacher in learning astronomy. *Science Education*, 91 (6), 948-966.

Shen, J. & Confrey, J. (2010). Justifying alternative models in learning the solar system: A case study on K-8 science teachers' understanding of frames of reference. *International Journal of Science Education*, 32 (1), 1-29.

Shen, J., Gibbons, P.C., Wiegers, J.F., & McMahon, A. (2007). Using research based assessment tools in professional development in current electricity. *Journal of Science Teacher Education*, 18 (3), 431-459.

Shen, J., & Linn, M.C. (2010). Connecting scientific explanations and everyday observations: A technology enhanced curriculum on modeling static electricity. Manuscript submitted for publication.

Yin, Y., Tomita, K. M., & Shavelson, R.J. (2008). Diagnosing and dealing with student misconceptions about "Sinking and Floating." *Science Scope*, 31(8), 34-39.

Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). A Comparison of two construct-a-concept-map science assessments: Created linking phrases and selected linking phrases. *Journal of Research in Science Teaching*, 42(2), 166-184.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ: Erlbaum.

# Using changes in framing to account for differences in a teacher's classroom behavior

Jennifer Evarts Lineback, Fred Goldberg, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182
lineback@sciences.sdsu.edu, fgoldber@sciences.sdsu.edu

**Abstract:** In this paper we account for how a 5th grade teacher changes in how she attends and responds to student thinking in terms of differences in her framing of the instructional situation. Coordinating verbal (i.e. speech, voice inflection, and word emphasis) and non-verbal (i.e. body orientation, eye gaze, and gesturing) behavioral data, we characterize three patterns that reappeared with stability throughout a fourteen-day, inquiry-based instructional module on the water cycle. We argue that each of these behavior patterns suggests a distinctive way the teacher frames her instruction. We label these frames as: (A) moving towards content objectives; (B) promoting a student-centered discussion; and (C) making sense of students' ideas. The episodic manner in which these frame shifts occur in the classroom suggests a way of conceptualizing a learning progression in teachers' ability to attend and respond to student thinking.

## Introduction

The following transcripts document two brief exchanges between a fifth grade teacher and her students that took place on successive days during a fourteen-day, inquiry-based, science unit on the water cycle. Exchange #1 is part of a classroom discussion where the teacher and students are trying to understand an idea for an experiment suggested by a student (Jasper). Exchange #2 is part of a classroom discussion where the teacher (Mrs. Miller) is trying to get students to come up with specific ideas about precipitation. Words in italics indicate words emphasized by the speaker. Non-verbal behaviors that coincide with talk are indicated in brackets.

Exchange #1: Day 13

**Jasper:** If you did that with a *different* drink, that would work.
**Mrs. Miller:** Now *why* would you want to use a different drink? Other than water?
**Jasper:** 'Cause then you have no clue if it's leaking or not. I mean *that* could be water from a leak or water from the condensation.
**Mrs. Miller:** Like, what would you want to freeze?
**Jasper:** You could use, like … like soda or something, and then water appeared in the *bag*, you know it's- And you, like, *drank* it and you found out it wasn't like, soda or something-
**Jamil:** Oh yeah. I get what you're trying to say-
**Mrs. Miller:** OK. Alright. Go ahead. Go ahead. Finish what you're saying. (…)
**Jamil:** I get what Jasper's saying. Like, if you use, like, a different *drink*, like let's say, *orange juice* or something or *soda*, um- If it was *leaking*, … if you thought it was leaking *water*, you would know if it was from the *condensation*, because the orange juice is *yellow*, it's not all like clear like *water*.

Exchange #2: Day 14

**Mrs. Miller:** What s*teps*- notice that word right there- [points to the word "steps" written on the easel] What steps might be involved in making each type of precipitation? … And does somebody have a thought? Cam? (…)
**Cam:** First, rain is collected by the ocean, and the cloud carries it until it can't carry it anymore. So, it drops- so it- so when it can't carry it anymore, it drops it. It holds it until it's completely- it- it completely can't hold it anymore, at all. And, then it just falls.
**Mrs. Miller:** So, (…) does that define precipitation? Does what you say really tell me what *these* four [points to the types of precipitation] are? Or does it tell me what *that* [points to the word "precipitation"] is? What do you guys think? Is Cam telling me what *this* is, or is he telling me what *these* are?
**Jack:** He's telling you what rain is.
**Mrs. Miller:** *Really*?

Although the context for both teacher/student exchanges is the same (i.e. the water cycle), the verbal information gleaned from the transcripts suggests that the interactions taking place on these two days are substantively different. If these transcripts are analyzed along with the co-occurring extra-verbal behavior, such as the participants' body orientation, eye gaze, and gestures, the case is strengthened that something dramatically different is taking place in these exchanges. How can researchers best make sense of this shift in the dynamics of the classroom, particularly with respect to the teacher's practice? We propose that the construct of *framing* is a potentially useful framework to help answer this question.

## Framing as a Framework

Educational researchers have adopted the theoretical construct of framing in recent years to help account for the dynamic nature of students' and teachers' reasoning during instruction. A concept that originated in the disciplines of anthropology, sociology, and linguistics, framing refers to an interpretation of how an individual makes sense of "What's going on here?" (Bateson, 1972; Goffman, 1986; Hammer, Elby, Scherr, & Redish, 2005; Tannen, 1993). Framing is considered to be a mechanistic, cognitive process, serving to activate sets of locally coherent epistemological, conceptual, and social resources and enabling individuals to effectively navigate their current situation (Hammer et. al, 2005; Redish, 2004). An individual's frame, and the resources that are thereby activated, informs the set of expectations that he or she has about a situation, which is necessarily grounded in his or her previous experience with similar types of situations.

In the context of education, framing influences both what individuals attend to within instructional situations and how they think to respond in the moment. For example, a teacher's framing of instruction influences how and to what she attends in the classroom, as well as how she thinks to respond (Levin, Hammer, & Coffey, 2009). How students frame instructional activities helps determine the ways they interact with one another, their teacher, and the available learning materials (Hammer et. al, 2005; Scherr & Hammer, 2009). In an empirical study, Scherr and Hammer (2009) characterized distinct sets of co-occurring behavioral clusters that suggested that college students were framing their group activities in different ways at different points in time. Shifts *between* these frames were observed within single physics sessions, indicating that students did not maintain one consistent way of making sense of and approaching their learning activities. By considering frames as a theoretical framework, researchers may be able to better make sense of changes in students' and teachers' behaviors and posit *why* they behave as they do.

It has been previously suggested that when teachers attend and respond to the substance of students' ideas, the students are more likely to genuinely engage with the material under investigation and advance in their learning (Black and Wiliam, 1998; NRC, 2001). With respect to developing students' inquiry practices, a teacher's focused attention and response can, therefore, help students refine and develop their questions, investigations, and explanations surrounding specific natural phenomena (Levin, 2008; NRC, 2000). If attending and responding is a key component of promoting student inquiry, then it is critical for researchers to be able to account for how and why teachers attend and respond to the things that they do. Framing provides a vehicle for us to do just that. Characterizing the ways a teacher frames instruction enables us to make sense of how she attends and responds to students in the moment and explore how her attention and responses change longitudinally.

The objective of this paper is to present behavioral evidence that suggests different ways a single teacher frames her inquiry-based instruction within a unit on the water cycle. This work lays the foundation for research that investigates the frequency by which she shifts between frames, how this frequency changes over time, and how such changes may indicate progress in learning to attend and respond to student thinking. Furthermore, this research enables us to determine whether similar frames and shifting patterns are observed in additional teachers.

## Methods

The setting for this study was a fifth grade teacher's classroom during her implementation of a fourteen-day "module" on the water cycle. The module represented one of several developed by our research group designed to provide teachers with a context with which to explore and promote student inquiry practices (1). To accomplish the goal of encouraging practices such as making-sense of and reasoning about scientific phenomena, proposing plausible, mechanistic explanations of those phenomena, and collecting evidence to support or refute these explanations, the water module begins with the teacher posing an open-ended question to her students:

> Suppose that one night it rains. When you arrive at school, you notice that there are puddles of rainwater in the parking lot. When you go home, you notice that the puddles are gone. What happened to the rainwater?

The students are encouraged to suggest possible answers and consider reasons *why* these answers are plausible. While the minimal curricular materials provide teachers with potential student responses and possible ideas for follow-up discussion topics, the teachers are encouraged to base their future instructional moves solely on the nature of the students' ideas as they emerge during the discussion itself. Hence, the teacher may choose to go in a variety of different directions in this and subsequent class sessions depending on the responses of the students.

A single teacher, Mrs. Miller, was observed in her classroom throughout the duration of the water module in the spring of the 2008-9 academic year. Mrs. Miller is a national-board certified teacher and has taught elementary school for over twenty-five years. She has served as a mentor teacher for novices and has identified herself as a teacher comfortable with teaching science. Prior to the module implementation, Mrs. Miller participated in inquiry-based professional development (PD) activities organized by our research group for just under one year. Our PD sessions provide opportunities for in-service elementary and middle school teachers to practice attending and responding to student thinking, with the explicit intent of helping them develop their ability to promote rich student scientific inquiry in the classroom.

Mrs. Miller's class sessions during the module were video recorded in entirety. This video served as the primary source of data for analysis, supplemented by field notes taken by the first author both during class and during debriefing interviews with the teacher that occurred shortly after each instructional session. Since the focus of the research was on *teacher* framing, portions of the classes that centered on student group work and experimentation, which tended to involve minimal teacher participation and to which we had less access, were excluded from the data analyses. Methods of analysis were modeled after those used by Scherr and Hammer (2009), in that video data was evaluated for patterns that displayed specific sets of Mrs. Miller's verbal (i.e. speech, voice inflection, and word emphasis) and non-verbal (i.e. body orientation, eye gaze, and gesturing) behavior that seemed to consistently co-occur. Once such behavioral pattern sets were tentatively characterized, they were compared with additional segments of class video to determine whether they occurred with relative frequency and stability. From those that were established to be regularly occurring patterns, inferences were drawn regarding how these patterns might reflect Mrs. Miller's framing of the instructional session.

## Results

Three distinct and coherent patterns of Mrs. Miller's verbal and nonverbal behaviors seemed to emerge from the data, which we are calling patterns (A) "focusing on content," (B) "focusing on generating student interaction and discussion," and (C) "focusing on making sense of student ideas." These patterns seemed to occur with some regularity throughout the module, with some occurring more frequently and lasting for a longer duration than others. Interestingly, these patterns did not seem to appear in a specific chronology. For instance, pattern (A) was not only observed in the earliest class sessions of the module, nor was pattern (C) detected only in the final sessions. The data also indicated that Mrs. Miller's students were actively engaged in discussion throughout the module, no matter the pattern of behavior exhibited by Mrs. Miller. The types of interactions in which they engaged and the substance of the discussion that took place, however, did appear to shift in coordination with Mrs. Miller's behavior.

In the sections below, each of Mrs. Miller's three behavioral patterns is described in detail. Supporting evidence for these patterns is provided via portions of classroom transcript, digital stills taken from video, and field notes taken from the debrief interviews. These data provide the basis for inferences drawn regarding how Mrs. Miller was framing her instruction. The emerging behavioral patterns were considered to be stable in that they were consistent for *at least* several exchanges between discussion participants. It is worth noting that while there were occasions where a single behavioral pattern seemed to last for the entire class session, there were several occasions where Mrs. Miller shifted between different sets of behavior(s) at different points during a single session.

### Pattern (A): Focusing on content

Behavior associated with pattern (A) was observed regularly throughout the module and often lasted for large portions of the class session. Throughout these segments, Mrs. Miller regularly employed gestures emphasizing and/or enacting specific words or phrases, seemingly to help her students focus on particular concepts. These gestures often involved pointing out important terms or characters selectively inscribed on an easel (Fig. 1b) and/or drawing attention to specific content terms she employed in her verbal questioning. For example, in Figure 1c, Mrs. Miller held her hand up above her hand at a horizontal angle, emphasizing the idea of cloud height to her students.

With respect to teacher talk, Mrs. Miller tended to initiate these discussions by asking pointed questions in a direct, clear voice (e.g. line 1-2 in Exchange #2 shown in the introduction). These questions were often prefaced with "what is/are" and repeated at various intervals throughout the session to keep the students following a particular line of thinking. Her intonation during her speech, like her gestures, tended to emphasize content specific terms and phrases. Mrs. Miller seemed to pay close attention to students' terminology, as reflected in her repeating or emphasizing a particular term or phrase volunteered by a student and/or asking a student follow up questions regarding a term or phrase he or she used. Furthermore, Mrs. Miller selected certain pieces of her students' talk for display on the easel, seemingly choosing those phrases that advanced the discussion towards content goals.

Teacher/student interactions throughout these discussions tended to follow the traditional initiation, response, evaluation (IRE) format (Mehan, 1979).

Figure 1: Examples of Mrs. Miller's body orientation and gestural movements consistent with behavioral pattern (A): (a) Mrs. Miller creates a summary table that her students are to fill out; (b) Mrs. Miller gestures across her easel, directing their attention to the "steps" she is looking for; (c) Mrs. Miller emphasizes a particular descriptor she is looking for in her students' responses.

The following section of transcript displays an interaction between Mrs. Miller and her students taken from Day 2 of the module. The discussion illustrates behavior typical of pattern (A) and centers on the different types and locations of clouds.

**Tommy**: When you're at the, like the, um, *ocean*? To the *west*, in *early*- you get dark clouds. Usually, it would be *marine* layer, and it- um, most of it would be moving in, and then when it gets in enough, it will just come *into* a cloud. I don't know- [Mrs. Miller writes: "West early, dark clouds moving in."]

**Mrs. Miller**: But, ah- wh- are they *high* [moves hand horizontally above head, indicating "high"] or *low*? What's the *altitude*, I guess is what I'm asking. What's the *altitude* of the clouds?

**Tommy**: Well, when they're by the ocean, they're gonna be low. And then, when- When they're going on land, they're just going to drift a little higher and higher. [Mrs. Miller writes: "low" and "higher."]

**Mrs. Miller**: 'K, so they're gonna get *higher*. *Is* there someway we can connect how high they are with what they *look* like? (…)

**Cam**: Um, like, maybe when, like, there's storm rainy clouds, *way* high, maybe that means it just gonna drizzle or something. But, when they're *lower*, they're gonna- it's gonna, like, pour and rain really hard.

**Mrs. Miller**: Lower. Clouds. *Heavy* rain. *Higher* clouds. Drizzle. [Mrs. Miller writes on a new sheet of paper: "Lower clouds, heavy rain. Higher clouds, drizzle."] What does anyone else think about that? About *clouds* and their *altitude* as to how they look?

Notice in lines 1-4 that Tommy does not refer to the height of the clouds. Nevertheless, Mrs. Miller introduces the concept of altitude (line 6) and then goes on to emphasize specific terms in her own talk and in that of her students (e.g. *high*, *low,* and *altitude* in lines 5-6 and *higher* in line 9). She appears to be trying to get her students to recognize a relationship between clouds' appearance and their altitude. This is reinforced by Mrs. Miller inscribing what she identifies as the most pertinent words and phrases (e.g. "Lower clouds, heavy rain. Higher clouds, drizzle." in line 14) on the easel and repeatedly asking questions about a relationship between clouds and their location (e.g. "*Is* there someway we can connect how high they are with what they *look* like?" in lines 9-10). During her debriefing interview, Mrs. Miller's focus centered on the fact that her students' thoughts "were really scattered" during the discussion and they weren't "getting it." She seemed disappointed that they hadn't reached specific conclusions.

This pattern of behavior, exemplified by Mrs. Miller in both this segment and Exchange #2 in the introduction, suggests that she is framing her instruction in these types of discussions as *moving toward content objectives*, which we shall refer to as "frame (A)." Her direct line of questioning, reiteration of questions, emphasis on terminology and phrasing, and her later description of students "not getting it," all appear to be consistent with maintaining this type of content-driven frame. Rarely during these types of interactions would Mrs. Miller allow herself or her students to veer onto trains of thought that differed from this singular agenda.

### Pattern (B): Focusing on promoting student interaction and discussion

Similar to pattern (A), behavior associated with pattern (B) was observed throughout the module and lasted for large portions of the class session. In these discussions, Mrs. Miller neither directed her students' attention to an easel nor to her notes, as was commonly observed in pattern (A). In fact, she did not seem to direct the students' attention to herself at all. Rather, Mrs. Miller seemed to position herself, both verbally and non-verbally, as a discussion facilitator and "traffic cop," deciding who speaks next and who must wait. Her gestures seemed to be less emphatic or enactive than those observed in pattern (A) and were usually limited to pointing out the sequence of students next in line to contribute to the discussion (see Figure 2).

With respect to teacher talk, Mrs. Miller would generally initiate discussions with an open-ended question about some type of phenomenon (e.g. "Which clouds cause lightening?"). Subsequent to this initial question, however, her verbal contributions were generally limited to brief, non-committal "mmm hmmm's," and/or to call out the names of the students next in line to participate. Occasionally, she would insert a brief comment indicating that one student's contribution added to or responded in some way to a comment made earlier by a different student. She rarely evaluated or probed the students' ideas. Instead, Mrs. Miller seemed to expect that the students would comment on and/or ask questions about *each other's* contributions. These segments did not follow a single conceptual strand or delve particularly deeply into a single explanation or observation. Rather, students seemed to pursue several different trains of thought, dependent upon which ideas they decided to take up in the moment.


Figure 2: Examples of Mrs. Miller's body orientation and movements consistent with behavioral pattern (B): (a & b) Mrs. Miller points to the next student to speak; (c) Mrs. Miller faces a student as he contributes to the discussion.

The following transcript, taken from Day 3 of the module, is typical of interactions consistent with pattern (B). The focus of this session was on the causes and consequences of lightening. It is worth mentioning here that Mrs. Miller's behavior varied little during this entire sixty-minute session, reinforcing the idea that pattern (B) behavior is often stable for long segments of class time.

**Jack**: If you cut down *every* tree in the world, which isn't possible, but if you *did*, and you had nothing attracting lightening, it wouldn't be able to- um, um, some place twice. (4 sec pause.)
**Mrs. Miller:** Alright. [Points to Tommy.]
**Tommy**: Well, I was thinking that, if- if- *lightening* comes, is it just how much h- heat or electricity it gets? To touch the ground? Like, how many *volts* it collects? [Mrs. Miller writes in her notebook and looks back up at Tommy.] Or, (…) if a cloud has power, how much power does it get, compared to a *smaller* cloud, you know? If a *big* thundercloud, it might have a lot of it, and always strike the ground…
**Mrs. Miller**: Mmmm. Sam?
**Sam**: If you're in a *submarine* (…) Ah, not that that deep down, um, and lightening, um, went *down*, would it be able to, um, shock the submarine from underwater? [Mrs. Miller looks down at her notebook and up at Sam.]
**Harry**: If it hit the water [Mrs. Miller looks at Harry.], it would probably be able to send an electric current through it (…)
**Mrs. Miller**: Danny and then Len.

In this dialogue, Mrs. Miller's words are limited to non-committal responses or calling out students' names (e.g. line 9 "Mmmm. Sam?"), and her gestures are constrained to pointing to students. The student ideas are on display and tend to dominate the conversation. This segment also shows that, as commonly observed in pattern (B), multiple student ideas are available at any one time, many of which are never taken up by other participants in the class (e.g. Jack's ideas in lines 1-2 and Tommy's ideas in lines 4-8). In the debriefing interview that followed the class session, Mrs. Miller commented on the large number of students participating in the discussion and the frequency of that participation. While she was able to point out a few emergent student ideas when prompted, her focus was more on the participatory nature of the discussion, rather than on the richness of the expressed ideas.

Mrs Miller's pattern of behavior during these segments of instruction appeared to sharply contrast to that of pattern (A), which tended to be much more teacher-directed. The verbal and non-verbal behaviors of Mrs. Miller depicted in pattern (B) interactions suggest that she is framing the instructional sessions (or portions thereof) as "making room for students to share ideas and reach their own conclusions about phenomena with minimal interference from the teacher," or *promoting a student-centered discussion* for short. We will refer to this as, "frame (B)." Mrs. Miller's limited talk and her propensity for calling out the names of students seems consistent with eliciting student ideas and encouraging student participation. Mrs. Miller seemed to be intent on making room for her students to express their thoughts and allowing the students' ideas to take the floor. Rarely did she try to control and/or direct the trajectory of the students' ideas. Her focus on participation during her debrief interview, rather than the substance of her students' ideas, seems to reinforce this interpretation.

## Pattern (C): Focusing on making-sense of student ideas

The set of behaviors associated with pattern (C) was observed less frequently and for shorter duration (usually less than 15 minutes) than those associated with patterns (A) and (B). Similar to that which was observed in pattern (B), Mrs. Miller's behavior made room for student ideas and allowed them to be objects of discussion. A substantive difference between (B) and (C), however, was Mrs. Miller's active role *during* that discussion. In these sessions, Mrs. Miller seemed to concentrate intently on the substance of the students' ideas, leaning towards the student as he or she spoke. She also actively employed gestures during these interactions, seemingly to enact the situation that the student was describing (see Figure 3a). On one occasion, she even spontaneously constructed a simple demonstration to further enact the student's idea being described (see Figure 3b). These non-verbal gestures are in contrast to those associated with pattern (A), where she tended to use hand and body movements to emphasize specific content terms or phrases in her own talk rather than those of her students.

**a**

**b**

Figure 3: Examples of Mrs. Miller's body orientation and gestural movements consistent with behavioral pattern (C): (a) Mrs. Miller uses arm gestures to enact a particular idea posited by a student (see line 14 in transcript below for co-occurring talk); (b) Mrs. Miller constructs a demonstration enacting a student's experimental design.

Similar to what was observed in behavioral pattern (A), Mrs. Miller tended to be more engaged verbally during these discussions. A major difference, however, is her engagement with the students' ideas, rather than her own. During these sessions, she actively asked questions of her students, emphasizing their words and probing more fully into their thoughts about the phenomena under discussion (e.g. "Now why would you want to use a different drink? Other than water?" line 2 in Exchange #1 listed in the introduction). In these discussions, Mrs. Miller seemed to act as a participant, similar to the other members of the class, working with the ideas the students proposed. At times, she attempted to repeat the students' ideas back to them, following such comments with inquiries as to whether she had "got it" or if she "was putting words in [their] mouth." Such questions indicate that she was checking in with her students to determine if she fully understood their ideas.

The following is a transcript that reflects an interaction typical of behavioral pattern (C). It was taken from a discussion that occurred on Day 12 of the module, which centered on a question posed by a student (Tommy) concerning a phenomenon that he had witnessed.

**Danny**: I just *know* I've seen it before. On the car windows-
**Mrs. Miller**: So more on a *car* window, though. [Looks at Tommy.] Is it the same idea as a car window? Or is it *different*? Are you thinking something different?
**Tommy**: It's almost the same concept except- except *that* is sideways, and I'm talking about, ah- up- you- ah- up, like, so as a straight area, so it doesn't have any *tilt*. You know? You have a window that's *this* way [Tommy tilts his right hand downward], you know the drop of water's gonna come *down* [Tommy indicates drops sliding down with his left hand]. And if it's *this* way [Tommy holds his right hand horizontally], it'll come toge- I don't know why, it just comes together, like, you know.
**Mrs. Miller**: So, you're *asking* if it's on the ceil- First, how's it get on the ceiling?
**Tommy**: Well, ah- a- I don't know, it was just thinking maybe if you splash water or something.
**Mrs. Miller**: Oh. If you splash water or something?
**Tommy**: Ah- ah- yeah.

**Mrs. Miller**: So, it [looks up at her hands, which are raised to the ceiling] comes together- [moves her hands together]
**Tommy**: And then it makes one drop of water and falls.
**Mrs. Miller**: Instead of coming just straight down. [brings both hands downward]
**Tommy**: Yeah- like, why does the water *move*? It doesn't have [Mrs. Miller looks up at right hand.] any reason to. [Mrs. Miller drops hand downward.]

Mrs. Miller's interactions with Tommy in this transcript clearly indicate that she is trying to understand the phenomenon that Tommy is trying to describe. She asks several pointed questions (e.g. "Is it the same idea as a car window? Or is it different?" in lines 2-3 and "How's it get on the ceiling?" in line 9) and employs gestures throughout the exchange. This use of gestures seems to serve two objectives: (1) to help herself visualize the phenomenon better, and (2) to ask Tommy if she has grasped his idea correctly. In total, the verbal and non-verbal behaviors of Mrs. Miller displayed during pattern C suggest that she is framing this portion of the instructional session as *making sense of student ideas*, which we refer to as, "frame (C)." Her active participation, probing questions, and use of gestures seems consistent with trying to understand her students' thoughts and further the development of these thoughts. Overall, these appeared to be richer discussions than those that occurred when Mrs. Miller employed frames (A) and (B) in that these conversations tended to extend and develop a single student's idea to its conclusion rather than following a series of different ideas superficially.

## Discussion

Mrs. Miller is a teacher who has historically taught science by focusing on reaching content objectives. Data collected from Mrs. Miller's classroom early in the 2008-9 academic year (not reported here) showed a clear tendency for Mrs. Miller to focus her attention on institutionalized content standards. After participating in PD sessions focused on attending to student thinking for one year, however, Mrs. Miller's verbal and non-verbal behavior in the classroom indicated that she no longer framed her instruction solely in terms of reaching content objectives. Instead, our findings show that while Mrs. Miller did demonstrate behavior that indicated that she was framing instruction as *reaching content objectives* in some situations, she also demonstrated that she framed her instruction as *promoting a student-centered discussion* and *making sense of students' ideas.* The employment of different frames, rather than a single frame, suggests that she is shifting in her approach to teaching science and facilitating scientific inquiry.

It might be expected that an experienced teacher, such as Mrs. Miller, would move between frames in a stable, predictable manner. This was certainly not the case. She did not appear to shift from one instructional frame to another sequentially, such as framing the initial module sessions as *reaching content objectives* (A), framing subsequent sessions as *promoting a student-centered discussion* (B), and framing the final module sessions as *making sense of students' ideas* (C). Nor did she seem to shift between frames in a cyclic fashion, moving from frame (A) to (B) to (C) rapidly, only to instantiate a similar cycle in later instructional sessions. Instead, Mrs. Miller tended to shift between frames sporadically, framing instruction differently on successive days and occasionally appearing to shift between frames within single instructional sessions. For example, Exchange #1 took place on Day 13 and provided behavioral evidence that Mrs. Miller was framing instruction as *making sense of students' ideas.* Exchange #2 took place the very next day (Day 14) and provided behavioral evidence that Mrs. Miller was framing instruction as *reaching content objectives*. Furthermore, Mrs. Miller displayed behavioral evidence (although not included here) that she was framing instruction as *promoting a student-centered discussion* just twenty minutes prior to the Exchange #2 segment.

Mrs. Miller once again participated in our research project during the 2009-10 academic year. Her second implementation of the water module provided us with the opportunity to explore how her framing of instruction may have changed from her first implementation. Preliminary behavioral analysis indicates that Mrs. Miller employs frame (A) considerably less frequently, whereas evidence of frames (B) and (C) seems to be more prevalent than was observed during her first implementation. Furthermore, while instances of frame (C) continue to be fewer in number and of shorter duration than those of frame (B), Mrs. Miller does appear to be shifting between these two frames more frequently *within* session than was observed during her first implementation of the module. Although these data are admittedly tentative, fewer instances of (A) (e.g. focusing on content only) and shifting from (B) to (C) more regularly (e.g. engaging with student ideas *in-the-moment* during student discussions) suggests that Mrs. Miller is progressing with respect to attending and responding to the substance of her students' thinking. Her debrief interviews from this fall seem to corroborate this interpretation, for her comments seem to be focused primarily on her students' ideas and how these ideas can promote additional explanation and discussion about phenomena. This focus contrasts with last year, when she tended to comment on whether her students did or did not achieve content goals and/or to what degree her students were participating in discussion.

Part of the broader scope of our research project is to begin to conceptualize a teacher learning progression in the promotion of scientific inquiry. We expect that such a learning progression necessarily involves a dimension in how teachers attend and respond to the substance of their students' ideas. Linear descriptions of learning progressions tend to characterize learning as happening in stable, plateau-like stages representing the increasing degrees of sophistication (e.g. Thompson, Braaten, & Windschitl, 2009). The dynamic, episodic nature of Mrs. Miller's observed behavioral patterns, however, suggests that it might be less meaningful to describe longitudinal progress in terms of successive stages or levels. While it is arguable that frame (C) is more sophisticated than frame (A) regarding a teacher's consideration of the substance of her students' thinking, it is questionable as to whether it is desirable (or feasible) for a teacher to only frame instruction this manner. Allowing students to first share their ideas and reasoning in a student-centered discussion might be a necessary foundation for engaging, probing, and working with their ideas. Characterizing *how* a teacher shifts in her framing of instruction, as well as the *frequency* in which she shifts between frames, might serve to be a more productive way to think about a learning progression in attending to and responding to student thinking. Hence, a learning progression in this dimension might consider the degree to which a teacher is able to move flexibility between frames *in-the-moment* as different students' ideas emerge through discussion.

From a research perspective, framing appears to be a useful construct to help make sense of the ways Mrs. Miller attends and responds to her students thinking. Such a construct allows us to move from just describing observable classroom activity to positing reasons for *why* she might be attending and responding in specific ways at different times during instruction. Furthermore, thinking of framing in terms of activation of specific resources provides a means to consider the possible explanatory mechanisms underlying that activity.

## Endnotes

(1) The work described in this paper was supported by National Science Foundation grant 0732233, "Learning Progressions for Scientific Inquiry: A Model Implementation in the Context of Energy."

## References

Bateson, G. (1972). A Theory of Play and Fantasy (Originally published 1954). In G. Bateson (Ed.), *Steps to an ecology of mind; collected essays in anthropology, psychiatry, evolution, and epistemology* (pp. xxvii, 545 p.) San Francisco: Chandler Pub. Co.

Black, P. & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan, 80,* 139-148.

Goffman, E. (1986). *Frame Analysis: An essay on the organization of experience*. New York: Northeastern University Press.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89-120). Greenwich, CT: Information Age Publishing.

Levin, D. M. (2008). *What secondary science teachers pay attention to in the classroom: Situating teaching in institutional and social systems.* (Doctoral dissertation, University of Maryland, 2008)

Levin, D. M., Hammer, D. & Coffey, J. E. (2009). Novice Teachers' attention to student thinking. *Journal of Teacher Education, 60*: 142-154.

NRC. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academy Press.

NRC. (2001). *Classroom assessment and the National Science Education Standards.* Washington, DC: Committee on Classroom Assessment and the National Science Education Standards.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom.* Cambridge, Mass.: Harvard University Press, 1979.

Redish, E. F. (2004). A theoretical framework for physics education research: Modeling student thinking. In E. Redish, C. Tarsitani & M. Vicentini (Eds.), *Proceedings of the Enrico Fermi Summer School, Course CLVI*: Italian Physical Society.

Scherr, R. E., & Hammer, D. (2009). Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics. *Cognition and Instruction, 27(2)*, 147-174.

Tannen, D. (1993). *Framing in Discourse*. New York: Oxford University Press.

Thompson, J., Braaten, M., Windschitl, M. (2009). *Learning progression as vision tools for advancing novice teachers' pedagogical performance.* Paper presented for LeaPS Conference, Iowa City, June 2009.

# Explaining across contrasting cases for deep understanding in science: An example using interactive simulations

Catherine C. Chase[1], Jonathan T. Shemwell, and Daniel L. Schwartz
Stanford University School of Education, 485 Lasuen Mall, Stanford, CA, 94305
cchase@stanford.edu, jshemwell@stanford.edu, danls@stanford.edu

Undergraduate students used a simulation to learn about electromagnetic flux. They were provided with three simulated cases that illustrate how changes in flux induce current in a coil. In the POE condition, students predicted, observed, and explained the outcome of each case, treating each case separately. In the GE condition, students were asked to produce a general explanation that would work for all three cases. A second factor crossed whether students had access to a numerical measurement tool. Effects of the measurement tool were less conclusive, but there was a strong effect of instructional method. Compared to POE students, GE students were better able to induce an underlying principle of electromagnetic flux during instruction and were better able to apply this principle to novel problems at post-test. Moreover, prior achievement predicted learning in the POE group, while students of all academic levels benefited equally from the GE condition.

Science education has learning goals that range from basic lab skills to beliefs about the sources of scientific knowledge. One enduring goal is for students to develop a deep understanding of phenomena so they can engage in the structure of scientific explanation. One way to characterize deep understanding is the capability and disposition to perceive and explain natural phenomena in terms of general principles. In this study, we show that deep understanding can depend critically on the way in which multiple instances of phenomena are presented to students and how students are instructed to explain those instances. The research is done in the context of undergraduate physics students learning about magnetic flux with a computer simulation.

It is common in science instruction to ask students to solve or conceptually explain a series of problems. One version of this approach is the Predict-Observe-Explain (POE) cycle (White & Gunstone, 1992). Students receive the set-up of an experiment and predict what will happen. They then observe the outcome and develop an explanation for why their prediction did or did not match the expected outcome. For POE and other sequenced formats, a series of questions or examples is carefully selected to help students instantiate a given core principle in multiple contexts, so that they develop a deeper, more abstract sense of the principle and learn the kinds of situations to which it applies. Formats such as POE are considered to be effective in part because they foster deep and often extended engagement with each new question or problem that students consider.

A risk of presenting students with a series of instances of a given principle is that students may treat each instance as unique and not grasp the underlying structure that links them together. Novices often have difficulty finding the underlying structure across instances that differ on the surface. In a classic study contrasting physics experts and novices, the experts categorized problems by their underlying concepts, such as energy conservation, whereas novices categorized them by their devices, such as springs or inclined planes (Chi, Feltovich, & Glaser, 1981). This encoding of surface instead of deep features would seem a likely pitfall of any pedagogy that engages students intensively with many instances of phenomena presented in series. For example, students doing POE might focus on the manipulation of a particular experiment, not noticing that it shares properties with a seemingly different manipulation. As a simple thought experiment, if a person adds a red solution to a beaker in one POE cycle to see what happens, and then adds a purple solution in the next cycle, it would be natural to treat red and purple as distinct manipulations, even though they are both cases of adding a color that contains red.

An alternative to instructional methods that have students work intensively with separate instances of phenomena is to have students explicitly consider multiple instances jointly. Contrasts among multiple, juxtaposed cases are known to support the induction of underlying structure if they differ on a few key dimensions (Beiderman & Shiffrar, 1987; Gibson & Gibson, 1955; Marton & Booth, 1997). Much like wine tasting, the process of comparing across cases helps people discern critical differentiating features that they might otherwise overlook (Bransford, Franks, Vye, & Sherwood, 1989). When students come to recognize invariant structure among cases with different surface features, they can schematize this invariant and more readily transfer this more general knowledge to new situations (Gick & Holyoak, 1980). Approaches to instruction that optimize contrasts have been successful in teaching statistics (Schwartz & Martin, 2004) and psychology (Schwartz & Bransford, 1998). O'kuma, Maloney & Hieggelke (2000) provide an example of this type instruction in science, wherein students are asked students to discover, apply, and explicitly state an underlying principle induced from a series of related cases.

However, merely engaging with contrasting cases does not automatically produce deep understanding. Our hypothesis was that *how* students were instructed to process multiple, related cases would be critical for determining whether they would notice and encode the underlying structure. In particular, without explicit prompting, students would be likely to treat each problem separately and miss the common underlying structure. This hypothesis is supported by research in the domain of analogical reasoning. For example, Loewenstein, Thompson, & Gentner, (1999) showed that simply asking students to process two cases presented together was not nearly as effective for schema abstraction as explicitly prompting them to compare the cases and describe their similarities. Likewise, Catrambone & Holyoak (1989) found that transfer of underlying principles was improved when students were explicitly asked to identify the deep features that were common to two analogs. In the current study, we furnished all students with a set of contrasting cases embodying a single underlying principle. We asked one group of students to provide a general explanation (GE condition) for all the cases, whereas the other group followed the more typical approach of predicting, observing, and explaining each case in turn (POE condition). Our hypothesis was that the GE approach would lead students to induce the underlying principle during the activities, which in turn, would lead to better understanding at post-test.

Scientific principles that explain natural phenomena often involve complex relationships that are difficult to conceptualize using everyday language. Mathematics can provide crucial vocabulary and syntax to support students' conceptual reasoning in the face of complexity. For example, researchers (Schwartz, Martin & Pfaffman, 2005) had younger students use POE with the balance scale (i.e., will the scale tip or stay balanced given weights on pegs at various distances from the fulcrum). They found that encouraging students to "invent math" to predict and explain the results led to much greater learning than encouraging students to explain in "words." Representing distances and weights as numbers enabled students to test possible relationships (i.e. the multiplicative relationship of weight and distance that balanced the scale) and make precise comparisons that were difficult to make using words.

The simulation used in the current study features a measurement tool that mathematizes the concept of magnetic field by expressing field intensity as numerical values separated into their vector components. We gave half of the students in the study access to this measurement tool and encouraged its use on the presumption that it would help them identify and reason more precisely about the contrasts and similarities across the three configurations.

In the current study, undergraduates in an introductory physics course learned about magnetic flux in the context of an interactive computer simulation. Simulations offer exciting new possibilities for science learning (de Jong, 2006;), but pedagogies for their use are new and evolving. Instructional design has focused on providing embedded scaffolds to support student inquiry in relatively open-ended tasks, so students produce optimal experimental runs of a simulation (e.g., de Jong, 2006). Rather than focus on inquiry, we took advantage of simulations' affordances for engaging students with a set of contrasting cases. To do this, we asked students to generate conceptual explanations from a series of three scenarios within a simulation. We expected that using the common POE model of instruction, which encourages intensive processing of individual cases, would lead students to see different scenarios in the simulation as unique, unrelated instances, like the red and purple solutions in our thought experiment. Therefore, we wanted to determine if the simple switch of asking students to find a general explanation for all the cases could overcome this likely problem and produce superior learning outcomes.

Thus, the design of the study was a 2 x 2, crossing the factors of General Explanation (GE) v. Predict, Observe, Explain (POE) by Measurement Tool (MT) v. No Measurement Tool (No-MT). We expected the GE group to gain a deeper understanding of magnetic flux because in comparing across cases, they would be more likely to induce the general principle. We also predicted that the GE-With Measurement Tool (GE-MT) condition would perform the best of all on our learning assessments, because the precision of mathematical representation would help them identify and reason about relevant contrasts.

## Methods

### Participants

Participants were 103 undergraduates in an introductory physics course on electricity and magnetism at a highly selective university. The study took place during one of the 50-min recitation sections associated with the course. Because many students needed to leave before the end of the section (often to get to another class), 23 students did not complete at least one of the four questions on the post-test, leaving us with complete data for only 80 students.

### Design

Sections were assigned intact but at random to the four treatments: GE-MT (n=20; 3 sections), GE-No-MT (n=25; 3 sections), POE-MT (n=20; 3 sections), and POE-No-MT (n=15; 2 sections). The unequal numbers of students in each condition were due to variations in section size (6-15 students) and the odd number of sections. The eleven different sections were taught by six different teaching assistants, and all but one of the teaching assistants taught two sections. To compensate for teacher effects, each teaching assistant taught one GE and one POE section. Both sections for a given teaching assistant were then randomly assigned to either the MT or No-MT condition.

## Procedure and Materials

During the lesson, students completed worksheets, which directed them to use the simulation to learn about magnetic flux in the context of electromagnetic induction (Faraday's Law). Students worked in groups of two to three with one laptop computer. All groups spent 25-30 minutes on the worksheets. Throughout the class, the teaching assistants moved from group to group, answering student questions. Teaching assistants were unaware of the study's hypotheses. At the end of class, students completed a brief post-test to assess learning outcomes.

The simulation (Figure 1) was a PhET interactive simulation (Wieman, Adams, & Perkins, 2008; available at http://phet.colorado.edu/simulations). This simulation allows students to move a magnet around the screen to light a bulb attached to a conducting coil. According to Faraday's Law, a changing magnetic flux in the coil will induce a voltage and light the bulb. The simulation represents the magnetic field as a sea of tiny test compass needles. Changes in field direction and strength are depicted by rotation of the needles and changes in their brightness.

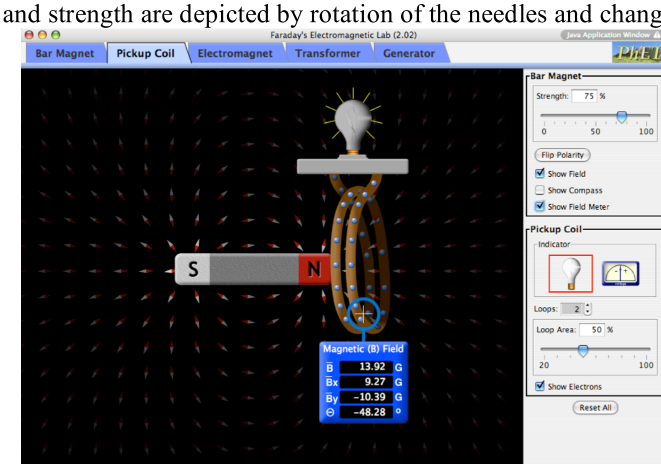



Figure 1. Phet Simulation. The magnet can be moved around the screen at varying speeds and positions to demonstrate how voltage is induced from magnetic field changes. The "field meter" measures field strength.

The worksheets presented all students with the same set of three cases (Figure 2). Two of the cases vary the magnet's position and one flips the magnet's polarity. A comparison of these cases reveals the invariant cause of voltage induction – a change in the component of the field within and perpendicular to the face of the coil. This translates to a change in the magnetic flux. The cases were designed to demonstrate three different manifestations of this underlying principle. Case A shows that a change in overall field strength can produce a voltage. Case C shows that a change in the field's direction can produce voltage. Case B illustrates that a change in field in the vertical direction does not produce voltage. Taken individually, it can appear that different kinds of changes are causing the voltage in each case. In Case A, the strength is changing; in Case C the direction is changing; in Case B there is almost no change in direction, and the change in strength is not particularly salient. Thus, a change in the field's strength or direction would seem sufficient to induce a voltage. But taken as a group, it is possible to induce the deep principle – that only a change in strength of the horizontal component of the field qualifies as a change in magnetic flux, which produces voltage.

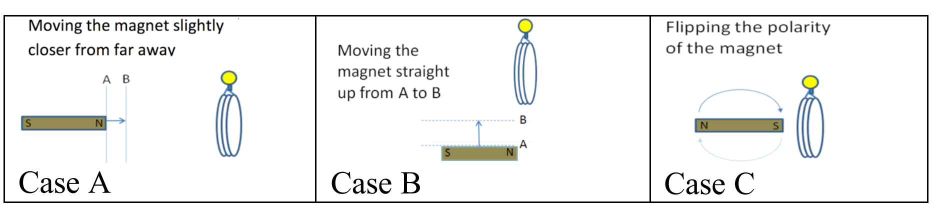



Figure 2. Cases. Students recreated these cases in the simulation and observed their effects.

## Description of Conditions

In the POE condition, for each of the three cases, students made predictions, observed what happened, and explained why. In the predict phase, the worksheets instructed students to make predictions for each of the three cases in Figure 2. Specifically, they had to predict what the light bulb would do in each situation and draw expected changes to the magnetic field. In the observe and explain phase, students used the simulation to test their predictions by recreating each of the three cases. For each case, there was a space on the worksheet for students to record "what the light did," describe the light's brightness, draw the changes that happened to the magnetic field, and "explain the change in the magnetic field that caused the bulb to light." Thus, the POE condition was unlike many POE cycles in that students worked with all three cases at the same time. This was done so that the POE condition would be comparable to the GE condition, which also worked with the three cases simultaneously.

The GE worksheets did not contain a prediction phase. Instead, GE students were told whether the bulb would light brightly or dimly, after which they observed the cases and worked to generate a single, unifying explanation that would work across them all. The worksheet contained an example general explanation for how three cases of objects of varying masses and volumes would sink to varying depths of a liquid (the example general explanation described density as a ratio of mass to volume, which determines sink "depth"). After looking over this example, students were instructed to open the simulation, produce each of the cases, draw and record observations of the magnetic field, and then write "a single general explanation that will address what the magnetic field must do for the bulb to light or not light in any given case."

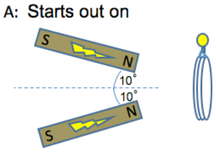
The field meter, an optional feature of the simulation (depicted in Figure 1) allowed users to take numerical measurements of the magnetic field. The field meter measured horizontal and vertical components of the field, the angle between the field vector and the vertical, and the overall magnetic field strength. Groups in the MT condition were given access to the field meter and told to use it to record horizontal and vertical components of field strength inside the coil. Studentsin the No-MT condition were told not to use the meter.

## Dependent Measures and Coding

During the last 10 minutes of the lesson, students individually completed a six-item test assessing their understanding of the vector (perpendicular component) contribution to changes in magnetic flux in the context of Faraday's Law. Two of the items were dropped from our analyses because they proved to be unreliable measures of student understanding of magnetic flux. Figure 3 shows an example post-test item.

Electric magnets A and B generate the same magnetic field as a regular bar magnet, but they can be switched on and off.

**Case 3**: Switch electric magnet A off at the same instant that you switch electric magnet B on. The field quickly fades from A and quickly builds up from B, so that the overall amount of field is held constant, but the direction of the field is changed.




Will the bulb light and if so, when? Why? Your explanation should discuss what happens to the magnetic field inside the coil.

*Answer: No, because even though the field changes direction, the amount of field perpendicular to the coil stays the same.*

Figure 3. Sample post-test item.

Post-test responses were coded for whether or not the deep structure (the vector component nature of flux) was discussed, using a 1-0 coding scheme. An answer with a score of 1 applied the principle that changes in magnetic flux depend on changes in the component of magnetic field perpendicular to the coil. We further subdivided the non-deep answers into two categories: shallow and vague. Shallow answers depended on surface features by referring to a change in the strength or direction of the magnetic field as the causal agent. Vague answers referred to a general change in magnetic field as the causal agent, without further specifying the type of change. In this shallow-vague coding scheme, shallow answers earned a score of 1, while vague answers earned a score of 0. Worksheet explanations were also coded along these two dimensions: deep structure and shallow-vague. All questions were coded by two primary coders. For each question, a random sample (20%) of the data was double-coded to achieve inter-rater reliability. Percent coder agreement ranged from 80-100% across questions.

# Results

## Equivalence of Groups

To check the equivalence of students across experimental conditions, we compared groups on prior achievement as measured by students' course midterm scores and found no significant differences. A factorial ANOVA on midterm scores crossed instructional method (GE or POE) with measurement tool (MT or No-MT). There were no differences in scores by instructional method, $M_{GE} = 27.3$, $SE_{GE} = 1.3$, $M_{POE} = 26.5$, $SE_{POE} = 1.5$, $F(1,73) = 0.17$, $p = .68$, nor was there an interaction of instructional method with measurement tool, $F(1,73) = 0.02$, $p = .90$. There was a near main effect of measurement tool, as the MT group had lower scores than the No-MT group, $M_{MT} = 25.3$, $SE_{MT} = 1.4$, $M_{No\text{-}MT} = 28.7$, $SE_{No\text{-}MT} = 1.4$, $F(1,73) = 3.30$, $p = .07$. However, this difference was in the opposite direction of experimental effects (described below).

## Post-Test Performance

Post-test measures revealed that the GE students developed a deeper understanding of the vector component nature of magnetic flux than POE students. There was a near-significant trend for MT students to outperform No-MT students, which suggests that using the field meter might also have helped students arrive at a deep understanding. Figure 4 depicts these patterns.

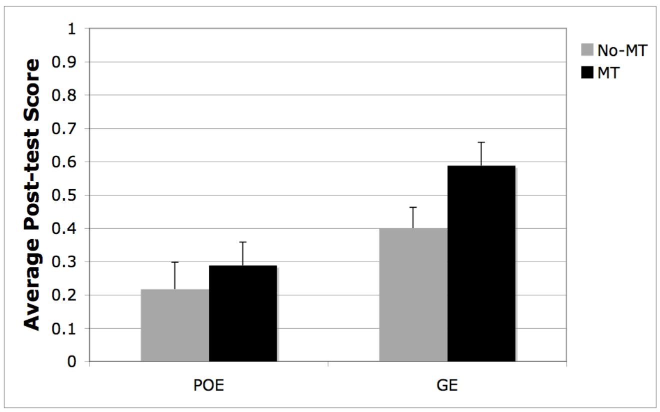



Figure 4. Average deep structure score across all post-test items, broken out by condition.

To test the effects of treatment on learning outcomes, a factorial ANOVA crossed method of instruction with measurement tool, using students' average deep structure score across all post-test items as the dependent variable. The ANOVA yielded a main effect for GE instruction, $F(1, 76) = 11.57$, $p = .001$, $d = 0.39$. There was also a near main effect of measurement tool, $F(1, 76) = 3.30$, $p = .07$. The interaction effect was not significant, $F(1, 76) = 0.67$, $p = .41$, though descriptively, the difference between MT and No-MT conditions was larger in the GE group.

## Effects of Prior Achievement

Pre-existing achievement levels predicted learning outcomes, but only for the POE condition. Correlations between post-test and course midterm scores were non-significant for the GE group, $r = 0.03$, $p = .83$, but moderate and significant for the POE group, $r = 0.39$, $p = .03$. Both MT, $r = 0.28$, $p = .09$, and No-MT, $r = 0.12$, $p = .46$, students' post-test scores were uncorrelated with achievement. The low correlations between post-test and midterm for the GE groups suggest that the positive effect of GE instruction acted independently of students' prior achievement levels. The opposite occurred in POE instruction, where high achievers learn more from the instruction.

## Worksheet Explanations

While working with the simulation, students in the GE condition wrote deep explanations on worksheets at a much higher rate than POE students (Table 1). This effect was pronounced. For the 80 students completing the experiment, only 1 out of 35 (2.9%) in the POE condition wrote a deep explanation compared with 14 out of 45 (31.1%) in the GE condition, $\chi^2$ (1, N = 80) = 1.03, $p = .001$. Measurement tool, in contrast, did not significantly affect worksheet performance, $\chi^2$ (1, N = 80) = 0.08, $p = .78$. So GE students were far more likely to induce the deep structure during the worksheet activity than POE students.

Table 1. Percentages of students who wrote deep explanations on worksheets (n deep/n total), by condition.

| | No-MT | MT | Total |
|---|---|---|---|
| **POE** | 0/15 (0.0%) | 1/20 (5.0%) | 1/35 (2.9%) |
| **GE** | 7/25 (28.0%) | 7/20 (35.0%) | 14/45 (31.1%) |
| **Total** | 7/40 (17.5%) | 8/40 (20.0%) | 15/80 (18.8%) |

## Relating Worksheet Explanations and Post-Test Performance

How students performed on worksheets predicted how they performed at post-test. A one-way ANOVA used worksheet explanation (deep or non-deep) as the independent variable and post-test score as the dependent measure. There was a substantial main effect of worksheet explanation, $F$ (1, 78) = 30.08, $p < .001$, $d = .62$. Students who noticed the deep structure during instruction (and wrote about it in their explanations) were far more likely to apply the deep structure to novel problem situations on the post-test. Only one POE student discussed the deep structure of the worksheet, so it is impossible to determine whether worksheet performance predicts performance equally for both instructional conditions. Nonetheless, the findings are clear. Worksheet performance strongly predicted post-test performance, and students in the GE group were far more likely to perform well on the worksheet.

## Descriptive Trends in Non-Deep Worksheet and Post-Test Responses

What were students' worksheet explanations and test responses, if not deep? Figure 5 shows the percentages of student worksheet explanations and post-test responses that were deep, shallow, and vague, broken out by condition. Overall, students across all four conditions gave similar proportions of vague responses, meaning that POE students did not simply neglect the task, at least not more than the GE students. However, the POE group had a higher percentage of shallow responses on both the worksheet and post-test, indicating that POE students tended to focus on locally salient aspects of the three cases (i.e. the field changed strength, or it changed direction). In comparison to MT students, No-MT students wrote a higher proportion of shallow explanations at test, though this pattern was far less prominent on worksheet performance. As is evident from the data already presented, the GE group generated a higher percentage of deep responses on both worksheets and tests.

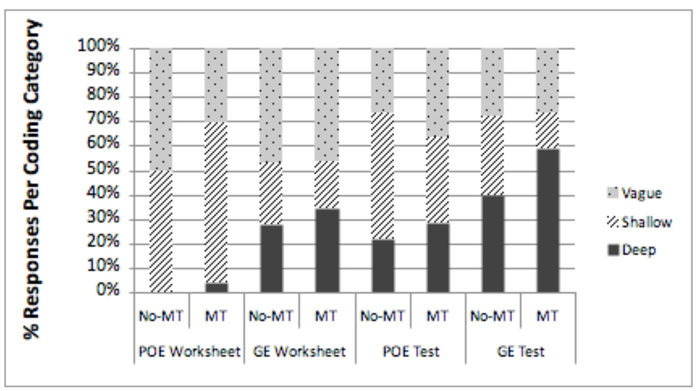



Figure 5. Percentages of deep, shallow, and vague responses on worksheets and post-tests.

## Mortality Threat

As described in the methods section, 23 students did not complete one or more of the assessment items because they left the experiment during the assessment phase. These students' worksheet and partial test performances are omitted from the analyses described above. To check for mortality effects, we calculated the percentage of students who answered in terms of deep structure for each post-test question and on worksheets, including all student data (Table 2). Importantly, the pattern of worksheet performance including all 103 students is identical to the pattern for the 80 students who completed all post-test items. At post-test, GE students gave deep responses at a higher rate than POE students on all four items. Also, with the exception of Item 3, MT students gave deep responses at a higher rate than No-MT students. In a final check, we examined differences in prior achievement by comparing leaving and remaining students' midterm scores. Scores in the two groups did not differ, $M_{leave} = 26.8$, $SE_{leave} = 1.5$, $M_{remain} = 27.0$, $SE_{remain} = 1.0$, $t(1,97) = 0.09$, $p = .93$, nor did they differ by condition. Given the equivalent prior achievement and the results shown in Table 2, and especially considering the strong relationship between worksheet and test performance, it is unlikely that results described above were caused by attrition.

Table 2. Percentages of students with deep worksheet explanations and post-test responses, broken out by condition and item. Includes all student data.

| | | **Worksheet** (N = 103) | **Item 1** (N = 102) | **Item 2** (N = 101) | **Item 3** (N = 90) | **Item 4** (N = 82) |
|---|---|---|---|---|---|---|
| **POE** | No MT | 0.0% | 0.0% | 5.6% | 55.6% | 22.2% |
| | MT | 4.4% | 13.0% | 17.4% | 47.8% | 26.1% |
| **GE** | No MT | 27.8% | 16.7% | 33.3% | 58.3% | 27.8% |
| | MT | 34.6% | 46.2% | 57.7% | 69.2% | 38.5% |

## Discussion

Three strands of evidence show the superiority of working to create single, unified explanations of several simulated cases of a phenomenon (GE condition) over predicting and explaining those cases separately (POE condition). First, GE students had higher post test scores, showing their deeper understanding of the vector component nature of magnetic flux. Second, better worksheet performance in the GE group strongly linked this deeper understanding to the process of constructing a single explanation that satisfied all cases. Finally, in contrast to post-test scores for POE, post-test scores for GE were uncorrelated with prior achievement, indicating that this pathway to deeper understanding was open to students of all levels.

It is worth emphasizing that the students in all the conditions were engineering students at a selective enrollment university who had received instruction on magnetic flux in their main lecture class, many of whom had extensive prior experience with procedural basics including vector decomposition. Therefore, the lesson we taught was far from an introduction to the topic but rather an opportunity for students to rediscover "old" ideas in a new context. These special conditions suggest caution when generalizing our findings to contexts where students know less about the topic beforehand. Nevertheless, the non-correlation of test scores with midterm scores indicates that GE-type instruction could be beneficial for students with a broad range of prior knowledge. Here we note that many students in the study "forgot" about the component nature flux when working with the simulation despite their exposure to this topic in their course. Apparently, recitation activities that drive students to the task of "rediscovering" what they have been told in lecture is a valuable – perhaps indispensable – addition to student learning.

How did producing a general explanation help students develop a deeper understanding? Our favored interpretation is that when students search for a single explanation that applies to all cases, they are seeking and inducing the invariant under transformation. They are asking, "What is it about the three simulated experiments that are the same?" When successful, they induce an invariant relation that gives a principled account of different surface transformations. By contrast, students who work on cases in isolation, as the POE students did, are more apt to notice and think in terms of surface features that differ in each case. Of course, simply seeking a general explanation is insufficient if the data one collects are poor. The three simulation scenarios were designed *a priori* to include optimal contrasts that would help highlight the invariant structure. The GE students took greater advantage of these optimal contrasts than the POE students.

In addition to treating the cases either separately or together, there were other differences between the GE and POE conditions. Students in the POE condition predicted and observed the outcome of each case while those in the GE condition observed each case but were told the outcome. We cannot rule out these other differences as potential causes of the learning effects. However, given the nature of students' worksheet explanations, the "cross-case vs. within-case" explanation seems most plausible.

These results do not warrant a negative judgment on POE or similar pedagogies (such as sequential problem solving) because our implementation was not meant to represent optimal use of POE. Rather, the results show that asking students to induce a general explanation across a set of cases has the important consequence of facilitating a deeper understanding. Moreover, the results show that it can be risky to expect that students will generalize across multiple cases when they are asked to give a separate answer or explanation for each one.

In addition to the GE effect, the data included a trend in which the measurement tool helped both GE and POE students learn the underlying principle, according to post-test data. However, MT students did not have better worksheet explanations than No-MT students. We conclude from the worksheet data that, in opposition to our prediction, the reasoning afforded by this additional mathematical representation did not help students induce the underlying structure of magnetic flux from the three cases. Assuming that mathematical representations can facilitate students' conceptual reasoning, it seems that those we made available to students via the measurement tool were not appropriate, or were not appropriately structured, for the task of inducing a general explanation. Additionally, it is puzzling that MT students came near to outperforming No-MT students at post-test even though their worksheet explanations were not different. One explanation for this disparity is that the post-test contained stronger cues for students' prior knowledge of the vector component nature of flux than the worksheet activity. These cues would likely have been of greater help to MT students as a result of their having worked explicitly with vectors via the measurement tool.

Pending further research, the current study provides two suggestions. One is for the design of simulations and another is for the design of instruction. Currently, the design of simulations suggests running multiple experiments, each with a single condition. This design pulls for something like a series of cases using a POE pedagogy, where students set parameters, make a prediction, observe, and explain. This makes it difficult to compare across multiple conditions, which is what real world experimenters often do. Perhaps simulations should allow the presentation of multiple cases simultaneously on the screen. This would permit the production of optimal contrasts, and with proper orientation, students could be guided to generate general explanations. A second learning from this study, confirming prior research and extending it to the context of conceptual learning in science, is that when students are presented with several instances of a phenomenon, they will not automatically search for the common structure that exists across them. Rather, students need to be pushed to do this. Science instruction that compels students to generate explanations that work for several different experiments or situations can help students construct deep understandings of general principles and the conditions to which they apply.

## Endnotes

(1) The first two authors contributed equally to this work and are listed alphabetically.

## References

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 640-645.

Bransford, J. D., Franks, J. J., Vye, N. J. & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 470-497). NY: Cambridge University Press.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1147-1156.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

De Jong, T. (2006). Technological advances in inquiry learning. *Science, 312* (5773), 532-533.

Dean, D.J., & Kuhn, D. (2006). Direct instruction vs. discovery: The long view. *Science Education, 91*, 384-397.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment. *Psychological Review, 62*, 32-51.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical Transfer. *Cognitive Psychology, 15*, 1-38.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review, 6*(4), 586-597.

O'Kuma, T., Maloney, D., & Hieggelke, C. (2000). *Ranking task exercises in physics*. Upper Saddle River, NJ: Prentice Hall.

Marton, F., & Booth, S. (1997). *Learning and awareness*. Mahwah, NJ: Erlbaum.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition & Instruction, 16*, 475-522.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. *Cognition & Instruction, 22*, 129-184.

Schwartz, D. L., Martin, T., & Pfaffman, J. (2005). How mathematics propels the development of physical knowledge. *Journal of Cognition and Development*, *6*, 65-88.

Wieman, C.E., Adams, W.K., & Perkins, K.K. (2008). PhET: Simulations that enhance learning. *Science*, *322*, 682-683.

White, R. & Gunstone, R. (1992). *Probing Understanding*. London: Falmer Press.

## Acknowledgements

We would like to thank Chaya Nanavati for her guidance and support with implementation, and the Physics Education Technology (PhET) Project for use of their simulation.

# Conceptual Confusion in the History Classroom

Chava Shane-Sagiv, The Hebrew University of Jerusalem, chava.shane-sagiv@mail.huji.ac.il

**Abstract**: This paper addresses the inherent difficulties of learning concepts in the history classroom—inherent, because unlike disciplines that use a specialized vocabulary for concepts, history draws on the linguistic resources of everyday life, and uses the same concepts in different historical contexts. Relying on an eight-month classroom ethnography in a Jerusalem high school, I show how seemingly straightforward concepts such as "anti-Semitism" and "Nationalism" become, in the transaction of classroom life, a site of confusion reflecting the complexity of 'school history': learning about the past while struggling to find a common ground between disciplinary terms, on the one hand, and student knowledge and interests on the other. A close investigation of this confusion calls attention to how the disciplinary—everyday life and schooling—may be at cross purposes, and how we can become more astute in how to deal with this.

## Major issue addressed

In this paper I explore the challenges students face when learning concepts in the history classroom. Learning concepts and conceptual change have been dominant topics in science and math education for the past few decades but are rarely discussed in history education (Hallden, 1997). While biology concepts such as photosynthesis have clear definitions within the discipline and such concepts are generally used only in a biology context and not daily life (excluding metaphors) questions like "What is a historical concept?" or "What is the definition of a concept used in history such as 'fascism'?" seem less clear. Certainly many definitions of fascism can be found in textbooks and elsewhere, but understanding that twentieth-century fascism in Italy is not the same as a Texas congressman calling his President a "fascist dictator" (July 17, 2009 [LPAC]) is a challenge unique to the discipline of history.

The whole notion of generalizing in history, which lies at the core of conceptualizing, is viewed by some philosophers as ahistorical (Mink, 1987). Seen through this approach, history is about the concrete and particular rather than something abstract and general. But while the use of concepts in history may at times cloud more than it clarifies, teachers and students are discussing history concepts, both directly and indirectly, in classrooms around the world, and textbooks and tests are addressing them more frequently. This paper explores the confusion caused by such concepts as anti-Semitism and nationalism in a high school history classroom, and argues that while the use of concepts in the classroom is unavoidable, it presents difficulties for understanding both the history topics being discussed and the concepts themselves.

Consider the following excerpt from a high school history classroom in Jerusalem:

Ms. Stern: Today we will be talking about the Jews in Russia in the nineteenth century —In previous lessons we talked about the Jews in Western Europe and their acceptance and rejection, and we called this rejection anti-Semitism, but we didn't explain the term. What is anti-Semitism?
A few Students: Anti--?
Ms. Stern: "Anti" means against, and "Semitism"--?
Same students: Uh . . .
Ms. Stern: Shem was one of Noah's sons. What's the connection between him and the Jews?
Moti: The Jews came from Shem.
Ms. Stern: The Semites were a group of nations that originated from "Shem," including the Jews, Arabs, Aramaeans, and nations from all over the Middle East that had a common biological origin.
Tamar: So it's not just a term for the Jews?
Ms. Stern: No, it doesn't just refer to the Jews, but it eventually became a term that only applies to Jews.
Male student: What do you mean by "Semitic Nations?"
Ms Stern: The Semitic nations are a group of nations that originated from Shem.---
Shimi: So how can Arabs be anti-Semites?
Ms. Stern: Because of a lot of different things that happened over a long period of time. Basically we understand that the term today refers to the Jews, even though "the Semitic nations" include the Arabs as well.
Male Student: Shem?
Ms. Stern: *[slightly impatient]* Yes, from "Shem"--

There are clearly communicative aspects in this discussion that I will address; however, I would like to focus on the attempts to transact disciplinary knowledge. In five minutes of discussion, the following topics were mentioned: Jews, Russia, the nineteenth century, Western Europe, anti-Semitism, Semitic nations, Shem, Noah, Middle East, Arabs. We move from Eastern Europe to Western Europe to the Middle East, and from the nineteenth century to the time of the Bible in seconds, as the teacher tries to define anti-Semitism. She wants her students to acquire a basic definition of the concept and learn what it means etymologically while at the same time how to use it in the context of their history curriculum. While for the teacher the literal definition of the term is important as part of understanding the narrative presented in class (actually a multilevel narrative involving different kinds of hatred against Jews as well as a specific narrative of Jews in nineteenth- century Europe), the students when presented with this literal definition seem lost. The definition takes them far from the narratives the teacher would like to discuss, and they "cling" to the parts they don't understand or find intriguing, like the word "Semites," which makes no sense to some of them or the reference to Arabs, which they find confusing given the context.

The discussion quoted above exhibits communicative aspects across different disciplines, aspects that have been examined in research on classroom life (Jackson, 1968), classroom strategies (Holt, 1967), and classroom discourse (Cazden, 2001), for example, the teacher hoping her students would grasp what she's trying to say so that she can move on to another topic; the students trying to take clear notes in class so that they can do their homework and pass a test. While all of these aspects represent at once ways of passing along knowledge and power struggles in the classroom (Lemke, 1990), reasoning within the field of history has unique aspects which this paper will explore.

The teacher's attempt to give a general definition of the term "anti-Semitism" runs counter to several major forces at work in teaching and learning history. The first is students' experience with the concept in their own lives; the second is how different historical events, and the detailed narratives that form them, are unique, as opposed to generalizing them across time and space. In other words, students growing up in Jewish-Israeli culture are quite familiar with the term "anti-Semitism"; they bring into the classroom their own understanding of a modern "hatred of the Jews based on their Jewishness" which is so often mentioned in the news and in their education. Their own understanding is challenged by past historical events that share the same conceptual term but come from a period unknown to them, as well as by the term's literal definition, which some students have never before considered.

The confusion these students face when learning about a term like anti-Semitism (which is considered a concept in the curriculum) cannot be viewed as just the result of poor teaching or bad curriculum (though these can always be improved). Rather, it calls attention to the multiple linguistic layers and clashes between professional discipline, school goals, and student experience and interests, aspects which are apparent in every history classroom. While historians have tried to define anti-Semitism and filled countless books with their struggle to do so (Langmuir, 1990), the term's definition occupies only a few lines in a textbook and a few minutes of class time. This is not likely to change; the real question is how to manage it.

Learning history involves complex cognitive processes (Leinhardt, Stainton, & Virji, 1994). Some scholars have even argued that it is unnatural to think historically (Wineburg, 2001). Understanding the past can become even more complicated and charged in the classroom, given curricula and textbooks that all have a say in which topics to cover (or not) (Nash, Crabtree, & Dunn, 1997), in terminology that doesn't always explain topics well (McKeown & Beck, 1990), and where teachers and students bring to bear different historical knowledge and aims (Hallden, 1994 ). All of these factors play out in learning concepts in the history classroom. One of my goals is to show how the difficulties that arise in using concepts in the history classroom reflect the challenges of 'schooling history'—or any discipline, for that matter—challenges that must be resolved before they can be incorporated them into history studies.

This study investigates the challenges that concepts bring to the history classroom while addressing the following questions:

- What are 'concepts' in the history classroom?
- How do cognitive aspects involved in learning about the past (Lee & Ashby, 2000; Shemilt, 1983) and sociocultural aspects of studying culturally charged (Barton & McCully, 2005) and identity-charged (Seixas, 1994) historical topics in school play out in the learning of concepts?
- What kind of changes can be found in students' explanations of concepts once they have encountered them?

## Potential significance of the work

This work claims that, first and foremost, we need to recognize the disjunction between how concepts are used in everyday life and in disciplinary history, and then understand the fundamental problem of generalizing within the discipline of history. This may be researched not only through experimental work but in daily classroom exchanges. It asks us to explore the epistemological and linguistic landscape of disciplinary reasoning in today's classroom. Second, it points out that the remedy occasionally cited in research on disciplinary reasoning in the classroom, that difficulties in disciplinary understanding can be solved by changing the words used by textbooks or teachers, is too simplistic. Challenges such as learning historical concepts cannot be easily overcome; oftentimes the best that a teacher can do is manage them.

## Theoretical and methodological approaches pursued

Data for this paper were drawn from a larger research program of a tenth-grade history classroom in Jerusalem, Israel. This research was done in the tradition of classroom ethnography (Schweber, 2004) focusing on disciplinary understanding (Wineburg & Wilson, 1988) and more specifically on the "transaction of subject matter in speech" in a history classroom with 'typical instruction.' During the course of an academic school year I observed and took notes and audio recorded all 42 history lessons across eight months of instruction. Recognizing that many forces are present in classroom life that can regulate as well as silence certain voices, the research was enhanced by the use of a questionnaire, interviews, and other methods not traditionally associated with linguistic ethnographies in the classroom. This was a single-site case study (Yin, 1994) and as such, not intended to be generalized to other populations, but rather meant to illuminate and theorize phenomena that occur in other settings.

Mixed methods were used both in collecting data and in analyzing them. I collected multiple sources of data that could shed light on how students learned the history content presented in class.

1."Pre" and "Post" Knowledge Questionnaire: A questionnaire was handed out to all the students in the class before instruction began, asking about the topics in the curriculum. The same questionnaire was given at the end of the school year.

2. Ethnography Observations: Over the course of the school year I observed the history class twice a week. I digitally tape-recorded all lessons while taking notes on student talk.

3. Semi-Structured Interviews: During the school year I conducted bimonthly interviews with a focus group of selected students whom I asked questions about the classroom discussions.

4.“Artifacts of practice”: I collected materials used by the teacher and students in class (what Ball and Cohen call "artifacts of practice" [(Ball & Cohen, 1999)] including the textbook, handouts, notebooks and tests.

The first stage of data analysis took place while collecting the data. Protocols of lessons and interviews were transcribed based on my notes and class recordings, and preliminary codes were generated, focusing on students' line of reasoning in the classroom discussions (Hallden, 1994b). All the students' questions and in-class remarks were analyzed, looking for patterns of responses to content knowledge, i.e., difficulty in understanding, engagement, personal identification, etc. The second stage of analyzing these patterns meant examining them in light of discourse theories on learning knowledge in classrooms, focusing on moment-to-moment interactions between the teacher and students and the students themselves (Lemke, 1990; Wortham, 2006), and drawing on theories of group and classroom argumentation, adapted to the exigencies of the history classroom (Pontecorvo, 1993; Schwarz, 2009).

## Major findings, conclusions, and implications

Another example of confusion that arose in defining and discussing a concept in the history classroom involved the term "nationalism." Consider this excerpt from a class discussion:

Ms. Stern: What is nationalism? We’re starting a new topic: nationalism. Please write this down [*as she writes on the board she is speaking slowly]*. Nationalism developed out of the French Revolution, which emphasized nationality—the nation, or the people—as the basis of the state and the regime *[finishes writing on the board]*. Before the French Revolution, who owned the country?

Students *[in unison]*: The king.

Ms. Stern: The king. The French Revolution basically highlights the idea that the basis for the state is not the king, but the state, the citizens of the state… What is this thing we call nationalism? Nationalism is a strong sense of belonging to the national group, the common denominator for all members of the nation and a political aspiration for independence of the nation. . . . Okay, let’s try to understand. Finish copying, and let’s try to understand what this nationalism thing is all about *[reads from the board the definition she just gave the students]:* A strong sense of belonging to the nation--what is the word that recurs?

Students *[in unison]*: Nation.

Ms. Stern: Nation. What is this thing called a nation? What is this group that people feel a part of?

Natan: A state.

Ms. Stern: No, the state relates to the ruler. What is a nation?

Shimi: A group of people.

Ms. Stern: A group of people who what?

Student: Who have something in common.

Ms. Stern: Something in common such as what? Let’s try to take this apart: A nation, a group of people. But there are many groups of people. We’re a group of people in this class. Are we a nation?

Students: Uh . . .

This excerpt illuminates aspects that surround the confusion caused by studying concepts in the history classroom. Unlike the case of anti-Semitism, here the teacher is trying to explain an abstract concept that is new and unfamiliar to the students. She tries to explain nationalism by walking them through a definition of the concept she has just given for the first time, drawing on their own knowledge and experience. Analyzed from the perspective of a transaction of content knowledge and ideas, and not as a criticism of her teaching methods, this example shows how students fail to follow the teacher's "line of reasoning" because they don't share her context or knowledge. More specifically, it shows how the attempt to explain a concept by exploring the students' experience (i.e., "We're a group of people in this class. Are we a nation?") doesn't help, because the notion of nationalism is too foreign and complicated—foreign in its origin, since most of the students were born and raised in the nation-state of Israel, making it difficult to imagine what it would be like to live in a pre-existing world, and extremely complicated because the country they live in is a nation state that doesn't exactly conform to the definition of nation and nationalism they are learning about (data not presented here shows how the latter caused a lot of confusion). This can be seen as an example of how cognitive challenges (grasping an abstract concept) intertwine with socio-cultural challenges of learning a concept in the history class (Jews in Israel can define their nationality as Jewish, Israeli or both)

Comparing pre- and post- student questionnaires in terms of concepts like anti-Semitism and nationalism showed that while students did learn new information, the concepts themselves remained fairly blurry ahistorical ideas. These findings reflect the other processes connected to learning history in school that are addressed in this work: students' prior knowledge, beliefs, and ideas about the concepts discussed in class, which can represent a clash between "official and unofficial history" (Tulviste & Wertsch, 1994), and the difference between schematic and specific narratives about the past (Wertsch, 2002). These students tended to stick to schematic narratives, which resemble ahistorical concepts (anti-Semitism equals hatred of the Jews); while mentioning the specific narratives they learned (as a modern phenomenon anti-Semitism differs from religious hatred, or that practiced in nineteenth-century Russia or France), they were not clear on how to connect the two.

This work suggests three major sources for the confusion caused by discussions of concepts in the history classroom: (1) the clash between professional and everyday meanings as used in history; (2) the clash between history as a discipline of the unique and the notion of generalization that underlies the social science disciplines; and (3) the clash between the endless and opened structure of historical narratives and the time-limited/stressful/test-oriented structure of classroom life.

Given the push toward using concepts in teaching history in Israel and around the world (especially with the rise of standard testing), a better understanding of the many factors involved in understanding general concepts in the historical context is crucial. These days little classroom time is spent learning historical narratives in detail; as a short version of complicated narratives, using concepts may seem an ideal solution. The problem comes when this kind of shortcut creates confusion both in the class and in the narratives students take away. This work examines this confusion and argues that it reflects the complex processes that are involved in both trying to understand the past and the teaching of it. As such, it cannot be solved either through teaching strategies or changing words in a textbook; rather, this is a challenge that can be managed only through careful attention and a concerted effort within the field of history education.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey-Bass.

Barton, K. C., & McCully, A. W. (2005). History, identity and the school curriculum in Northern Ireland: An empirical study of secondary students' ideas and perspectives. *Journal of Curriculum Studies, 37*(1), 85-116.

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning* (2nd ed.). Portsmouth, NH: Heinemann.

Hallden, O. (1994). On the paradox of understanding history in an educational setting. In G. Leinhardt, I. Beck, & C. Stainton (Eds.), *Teaching and Learning in History.* Hillsdale, NJ: Erlbaum.

Hallden, O. (1997). Conceptual change and the learning of history. *The International Journal of Educational Research, 27*(3), 201-210.

Holt, J. (1967). *How children learn.* New York: Pitman.

Jackson, P. W. (1968). *Life in classrooms.* New York: Holt, Rinehart and Winston, Inc.

Langmuir, G., I. (1990). *Toward a definition of anti-Semitism.* Berkeley: University of California Press.

Lee, P., & Ashby, R. (2000). Progression in historical understanding among students ages 7-14. In P. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching and learning History.* New York: New York University Press.

Leinhardt, G., Stainton, C., & Virji, S. M. (1994). A sense of history. *Educational Psychologist, 29*(2), 79-88.

Lemke, J. L. (1990). *Talking Science.* Norwood, NJ: Ablex

McKeown, M. B., & Beck, I. L. (1990). The assessment and characterization of young learners' knowledge of a topic in history. *American Educational Research Journal 27*(4), 688-726.

Mink, L. (1987). *Historical understanding.* Ithaca: Cornell University Press.

Nash, G., Crabtree, C., & Dunn, R. (1997). *History on trial: Culture wars and the teaching of the past* New York: Knopf.

Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. *Cognition and Instruction, 11*(3&4), 365-395.

Schwarz, B. B. (2009). Argumentation and learning. In N. Muller Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and education: Theoretical foundations and practices* (pp. 91-126). Dordrecht: Springer.

Schweber, S. (2004). *Making sense of the Holocaust: Lessons from classroom practice.* New York: Teachers College Press.

Seixas, P. (1994). Historical understanding among adolescents in a multicultural setting. *Curriculum Inquiry, 23*(3), 301-327.

Shemilt, D. (1983). The devil's locomotive. *History and Theory, 22*, 1-18.

Tulviste, P., & Wertsch, J. V. (1994). Official and unofficial histories: The case of Estonia. *Journal of Narrative and Life History, 4*(4), 311-329.

Wertsch, J. V. (2002). *Voices of collective remembering.* Cambridge, U.K.: Cambridge University Press.

Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past.* Philadelphia: Temple University Press.

Wineburg, S., & Wilson, S. M. (1988). Models of wisdom in the teaching of history. *Phi Delta Kappa, 70*, 50-58.

Wortham, S. (2006). *Learning identity: The joint emergence of social identification and academic learning.* Cambridge, U.K.: Cambridge University Press.

Yin, R. K. (1994). *Case study research: Design and methods.* Thousand Oaks, CA: Sage.

# "Let the Players Play!" & Other Earnest Remarks about Videogame Authorship

**Abstract:** This exploratory case study describes how middle school youth authored videogames at an urban summer camp using Kodu, a visual programming language and game creation tool. The study details how the youth took on authorship and gave and withdrew agency to their intended game players through a process of continuous revision. Connections to traditional literacy are made and future research on the transfer of literacy practices between informal and formal learning environments are suggested.

## Introduction—Youth as Producers

Youth production of videogames and other media has been largely ignored by researchers (Peppler & Kafai, 2007a), and despite calls by educational scholars for further research in this area (e.g., Alvermann, 2008), very little is known about how multimodal production overlaps with traditional literacy practices. Our exploratory case study begins to fill the gap in the literature by looking at how middle school youth used Kodu (2008), a free tile-based, animation and videogame-making tool, during an inner city summer program. It also attempts to link the worlds of traditional academic literacy with the unique world of videogame authorship through a constructionist environment.

There is over 25 years of research on constructionist tools (Papert, 1969; DiSessa & Abelson, 1986; Sykes, 2007; Resnick, Kafai & Maeda, 2003) and youth videogame making (Kafai, 1993; Habgood, Ainsworth, & Benford, 2005; Perciles, 2007; Kelleher & Pausch, 2006; Howland, Good & Robertson, 2006; Robertson & Howells, 2008; Burn, 2008). The videogame studies have illustrated how production has a motivational effect, boosts self-esteem, engages participants in storytelling, and provides a gateway into early programming. Additionally, the research on constructionist tools has focused on learning and teaching practices, e.g., youth collaborative and creative processes (Peppler & Kafai, 2007a, 2007c), technology literacy and fluency (Cooper, Dann & Pausch, 2003; Resnick, Kafai & Maeda 2003), media literacy (Peppler & Kafai, 2007b), and instructional approaches (Moskal, Lurie & Cooper, 2000; Rusk, Resnick, Berg & Pezalla-Granland, 2008). Resnick and Silverman (2005) have also suggested that interactions with programming languages through constructionist tools "provides [users] with experience in using and manipulating formal systems" which may be transferred to "many other domains (from mathematics to grammar to law)." Other researchers have echoed such sentiments (see Gee, 2003; Ito, Horst, boyd, Her-Stephenson, Lang, Pascoe & Robinson, 2008). The shape of this transfer between subject area domains, however, has remained unexplored and unformulated.

Tackling the issue of transfer between domains, this exploratory case study devised simple research questions for an open and fruitful investigation of Kodu, a tile-based programming environment based on When/Do statements. These questions were: How did the youth engage with Kodu? And how did this engagement intersect with traditional generative literacy practices?

## Methods

An inner city community center in the Pacific Northwest was chosen for the study based on its geographic accessibility, computer lab availability, compatibility of the computers with Kodu specifications, and willingness of staff to support the program. Camp Kodu met three days a week over three weeks for 2.5 hours each session for a total of 22.5 hours of Kodu exposure. The youth, who attended Camp Kodu, might best be described as "geeking out" according to Ito's (2008) "genres of participation" since they wanted to be in Camp Kodu to build a project as they learned technology by interacting with adults and receiving feedback from peers. They were self-selected and learned about the program through signs around the community and word of mouth at the community center and at other institutions serving low income, urban and minority youth. In all, eleven youth between 6th and 8th grade signed up for the class. Six of the eleven had regular attendance and were able to complete a final project for the program. These youth emerged as the focus of the study (Patton, 2002). Of the six core attendees, three were African American (Ky, Zeb, and Thomas), one was white (Eli), one was second generation Chinese (Dale) and one was second generation Ethiopian (Ahmed). (All names are pseudonyms.) We acted as participant observer as researchers and instructional support for the youth. Since the unit of analysis for the study focused on the youths' game making and literacy practices and activities rather than on the interactions between the instructor/ researcher and the youth, we felt that we could interact with students freely as we followed their actions, but making sure not to direct their end products.

## Data Gathering and Coding

Observations were recorded and tagged for four areas: acts of collaboration, problem-solving, programming focus and storyline generation. Often tagging categories overlapped, e.g., collaboration and problem solving might both be tagged for an observational unit. Informal interviews during observation were also common.

These exchanges were recorded, parsed by thought unit, and coded by the same system as above. Ten to twenty minute think aloud protocols were also used as students were encouraged to talk while they composed with Kodu. The traditional think aloud format proved difficult for the youth given their task. Consequently, cognitive interviewing (Willis, 1999) was used to probe students in situ about why they made choices during programming. These sessions were parsed and coded by the same system as described previously. To round out our qualitative analysis, we analyzed the youths' games for code sophistication, clarity and simplicity. After the first round of coding was completed and reviewed, the theme of audience emerged prominently in the data, and subsequently, we added a theory driven coding structure based on Hammer's (2008) concepts of authority and agency, as it became clear these were of growing importance to the youth. We recognized that student uptake and use of the concepts would vary, perhaps because of the type of game they wanted to design or because of their comfort with Kodu or programming.

## Results and Analysis

The concepts of authorship proved fruitful for contextualizing our findings and illuminating the degree to which the youth crafted games with an eye toward audience. Traditional writing research has shown that revision and audience awareness are key components to an expert writer's compositional repertoire of skills (Bereiter & Scardamalia, 1987; Fitzgerald, 1987; Flower, Hayes, Carey, Shriver & Stratman, 1986; McCuthen, 1994, 1997). In this respect, authorship in videogame making is similar to traditional modes of writing; however, videogame authoring is murkier in its structure and more layered, as other New Literacies research has illustrated (see Hammer, 2008; Knobel & Lankshear, 2008; Miller, 2005). Consequently, our frame for analysis was most applicable theory to Hammer's concept of first, second, and third authorship in multiplayer role-playing gaming (2008). She theorized that "The primary text is that which outlines the rules and settings of the game in general. The secondary text uses material to create a specific situation, and the tertiary text is created as the characters encounter the situation in play" (p. 70). From an authorial standpoint, the first author develops the world, sets the rules, systems, and settings. They are considered the world builder who creates the sets and costumes. The second author, according to Hammer, constructs specific fictional situations (the scripts) within a pre-established imaginary world. The third author is the player who brings the story to life.

### 1st to 2nd Author

According to the above framework, the youth in Kodu occupied a space between the first and second author, suggesting that in constructionist environments authorship is more aptly described as degrees on a spectrum, which are influenced by the toolset established by the first author (in this case, the Kodu developers). Although Kodu has preset characters that are controlled through a series of preset tiles, the settings produce a myriad of environments, game genres, and character attributes. Such affordances fall more on the side of first authorship than second, but the degree to which these affordances are utilized is influenced by the user, as well as the genre the user is evoking. While the dynamics between what the first authors (Kodu makers) allow and what the second author (the youth) utilize is not the focus here, it should be noted that the youth were able to still be creative given the confines of the program by adjusting the settings of the world (gravitational pull, environmental colors/brightness, firing speed, etc.) and the settings of the characters (movement speed, level of springiness, etc.).

Additionally, the youth were also creative in their use of tiles (thus pressing on their abilities to evoke the first author through a problem scenario, not just via the tools that allowed setting changes). To demonstrate this, I challenged the youth to create a storm within a Kodu environment. There are no storm tiles within Kodu, but a number of tiles in combination can make a storm-like effect. This was an exercise in bricolage. For instance, one could clone a number of clouds that were colored gray or black and then add a yellow or orange glow to some, along with downward aimed wisps to illustrate the act of lightening. After this request, two students asked where to find the storm or rain tile while several others simply looked for the tile. I informed them that such tiles did not exist, and they had to figure out how to use the tool to create the effect. One student was overheard saying, "Oh, a storm. First I take a cloud and color it black to make it a storm," and Thomas went further and programmed the cloud to rain rocks when precipitation tiles were not found. These instances might be considered as the youth finding the bounds of their secondary authorship in connection to the first which constrained them to a degree.

As Hammer (2008) defined it, authorship is partially an issue of agency. Each author has agency, but the degree to which it is available is determined by the author above it. Unlike traditional texts which remain fixed in form but malleable in interpretation, secondary authors have the complicated task of formulating a fluid game-story. However, according to Hammer (2008), in order to compose they need to have various types of agency in certain amounts. Types of agency include: textual (freedom text code), narrative (ability to create one's own storyline), psychological (the feeling of authorial control) and cultural (others recognizing one's authorial control).

While in role-playing games, all these areas need to be activated in order for the secondary author to do her job, this may not be the case in Kodu, where it is likely that these areas of agency again run along a spectrum (as in authorship). For instance, cultural agency may not play as large a role in individual game-making as it does in the socially bound arena of role-playing. These points aside, however, it is important to point out that "Agency is not simply 'free-will' or 'being able to do anything,'" as Wardrip-Fruin argued (2009). "It is interacting with a system that suggests possibilities through the representation of a fictional world and the presentation of a set of materials for action." The notion of agency acting in a bounded system is sourced in the work of Mateas (2001) and reverberates in the work of Gee (2007), who positioned it as a key component of educating and one reason why videogames hold so much promise for learning.

## 2nd to 3rd Author

So this begs the question of how the youth acted as secondary authors in relation to their tertiary authors (their audience). What affordances of agency did they allow their game players and what were their contours of agency design?

In the bounded domain of game genre, some of the most potent examples of second-to-third author agency came to the foreground. The youth continually called upon a game genre on which to hang their hats—mostly racing and action. Although a variety of seed story games were available to the youth, they gravitated to the genres they played at home according to the surveys. Similar findings surfaced in Burn's research (2008) in which authorship emerged within the "proto-critical discourse" of gaming in general, Eli created a game called Left 3 Dead, a play off the zombie game genre in general, Left 4 Dead (2008) specifically. Interestingly, the original game served as a stencil of sorts by which to formulate a story, and Eli crafted a hybridized, remixed game (Knobel & Lankshear, 2008) using the DNA of Kodu with the game story of Left 4 Dead. Consequently, the genre and storyline not only deeply affected what he created, but also how the game was perceived and played by others. The general dynamics of this game were clear-cut for the player, and references to the original were deep.

In the original game, players controlled the four survivors of a zombie apocalypse, but Eli's game was transformed to three players, hence the name change. Zoey, Bill, and Francis were all taken from the original and used as characters in his version—though with a Kodu skin. Besides character names, however, Eli incorporated similar features and narrative plot lines to his game that were noticeable to his peers who were familiar with the game. Consider Eli's brief description of one part of the game he is creating, "They think everybody turned into zombies. And their friend is in this house and they have to rescue em. And then, that is how you win the game without dying because those zombies are trying to kill you by hitting you." Reflexively, these connections to the original game (which acted as another first author) bound the agency for the second and third author through replication of storyline and rules of game.

Dale used a similar strategy, designing a game around the Tower Defense genre, but he ran up against the player and genre expectation while creating his game, demonstrating how third author agency was violated. Dale had created a wonderful landscape—a maze made of steep canyons filled with water in which white and black Kodus started on opposite sides and tried to destroy each other's home base—a castle and factory respectively. We interviewed him about his game during the think aloud. After he showed in detail how his game worked, we probed for the role of the user.

| | |
|---|---|
| Researcher: | So what is the user controlling in this game? |
| Dale: | Ah, he is just watching. |
| Researcher: | Oh, he is just watching it so far? |
| Dale: | Yeah. |
| Researcher: | Is he going to be playing anything? |
| Dale: | No. |
| Researcher: | No? [long pause] Will he be playing anything? |
| Dale: | I don't think so. |

The exchange with the student surprised us a bit given the narrative foundation on which the student was building his game; Tower Defense definitely had an active user. We closely observed the student as he continued to develop the game and share it with others. At one point during the session, we had the campers play each other's games and give feedback. Zeb was particularly blunt in his assessment of Dale's game. He simply wrote, "Let the players play!" Clearly, he was feeling as if he had no agency and that his colleague (the second author) was impinging on his ability to actually be an active participant in the game. The player expectations rubbed too roughly against the player affordances, and Dale began to realize he was violating user expectation. He sat in his seat before making revisions, and stared blankly, "My game sucks." His game design was more attuned to an animated story, but in subsequent drafts, he created an environment that better aligned with player expectations—including player control. It is perhaps interesting to note that Dale scored well on standardized writing measures and did not define himself as a game player. These two factors suggest perhaps that during the initial stages of game development, he was connecting with more traditional, closed lines of authorship to express himself virtually and forgetting to leave a space for the third author to move.

These examples demonstrate how genre, knowledge of narrative history, and canonical pull impact the agency of the second and third author. These limits are not necessarily bad things, as stated earlier, since they serve as a function of coherency. As Nickerson suggested (1999) participants within an environment need structure in order to support their actions. However, it is when a second author limits the agency of the third in an unexpected way we see conflict. Participant agency is a balance between being bound and having freedom, which is not necessarily located in the powers of characters, but through them. This allows the third author to feel as if they have freedom without a sense of anomie. This idea is also reflected in game authoring literature. Wardrip-Fruin (2009) posited that "players come to games with assumption about the domain of play. To play successfully they must transition from their initial assumptions about this domain…to understand, often largely implicit, of how it is supported by the software model." If the transition is too sharp, novel or illogical according to the users past experience—for instance, not even allowing game play—the user will react.

Despite Dale's foibles with giving agency to the third author, there was a growing concern and attention given to the user by all the youth as they progressed through the camp. As with programming in general, they constantly pedaled between the Kodu program interface and the experience they were trying to create for the user. At first this act was the perpetual testing of the code without much thought about the player—perhaps acting as testers versus players. During the second of half of camp, however, and with more time devoted to sharing work, the youth became more interested in how and to what success their game was being played. For example, Ky was particularly attuned to audience and at first took pride in the fact that no one could win his game. With time and player feedback, however, he adjusted code to make the experience player friendly. He did this so much so that he would adjust the code and settings of his game depending on which camp peer was playing his game. When Ahmed won Ky's game on the first try, Ky increased the difficulty whenever Ahmed played his game—even if the first try was a fluke as they seemed to all agree. Such intense preoccupation with revising became a common practice for Ky and Eli. These two also spent the most time polishing their games in accordance with audience and what they considered to be challenging play.

Part of this revision included enhancement to the player's controller. Most of the youth began with long lines of code that allowed the user to control their characters in certain ways. However, Ahmed, Ky and Eli supplied the user with more than just movement and shooting ability. Ky for instance, gave the player the ability to glow blue, which was supposed to "blind" the enemies that were after him. Such a defensive tactic was not only clever, but also was in response to others suggesting that his game may be too difficult, hence also allowing the user greater agency to move within an environment without being destroyed.

As another example, Eli supplied the user with an array of options to act with agency. Although longer programming strands do not necessarily make for better programming, just as longer sentences don't make for better writing (consider the works of Hemmingway or Vonnegut), Eli's code set was telling, reflecting the type of creativity that was elicited from the storm-prompt. As an example, he used a programming feature (reloading) in a way that it was not explicitly intended. Additionally, his code offered a myriad of actions for the user through the game controller (some might think overuse), signaling his reaching out to the third author. First, he gave the player the ability to communicate status information (reloading) to the other two players in the game, as well as express emotions through text (DIE!!!). Interestingly, he also automated the calling of another character within the game (e.g., "Bill!!!") perhaps launching the story into narrative action. This may be considered an attribute assigned to user experience (the affect of the game) over third author agency (the ability to move within the game) since the player was not in control of this feature.

Additionally, Ahmed used text to communicate within his game, too. In contrast, however, he programmed text to appear when the lead character was moving on certain land types. "Careful! Oh Yeah! Keep Going! Stay on target! Heating Up!" The message acted both as a warning and as an encouragement, perhaps educating the third author of the constraints Ahmed had put within the game to make the play more challenging. This text might be considered an informative note to the third author suggesting the parameters of their agency.

The creation and adjustment of landscape was another way that the youth balanced agency for the third author. While considerably different from the broad scope of Hammer's (2008) definition of "framework agency" as it relates to role playing games, the crafting of landscape proved to afford players with degrees of agency in relation to the virtual physical world. The youth's continuous revision of the landscape signaled its importance as a game-play element. Thomas, for instance, spent several class sessions developing a racing game in which the landscape consisted of jagged mountains with steep valleys, but the landscape the path rested on was too extreme for his players to navigate without falling off and getting stuck. This was a real drag on agency for both characters, one of which was controlled by the player and other of which was automated. During one 20-minute think aloud (which was part of a much longer observation), we watched Thomas continually revise the path, rather than adjusting the characters' settings (such as speed). The session consisted of widening the path, changing the steepness of the path somewhat, adjusting the angle of the turn, raising the turns, and swapping path types. Between each change he ran the program to see if the characters would fall off, and they fell repeatedly. At one point, he said, “Poor little cycle.” This episode illustrated how the youth tried to give agency by tempering the effects of steep terrain, as he measured the game’s challenge level.

While Thomas attempted to tame the force of landscape as he retained its aesthetic, Ahmed totally revised his landscape to tame it for character and participant movement, “making it a little more even for both the enemy and the player,” and Ky used the landscape as part of strategic play. While discussing his landscape, Ky had a mind toward the player continuously.

| | |
|---|---|
| Ky: | It makes it look like it’s more intense because of all these rock spaces and stuff. It is going to make the player feel like he is in an enclosed area so when he runs this, [directs the players character] so like when he goes this way, he is going to be trapped when he turns around there is going to be a dead end with all the zombie Kodu. |
| Researcher: | So the landscape is influencing the story you are telling? |
| Ky: | Oh yeah, and I forgot there is a safe space. |
| Researcher: | Oh really? |
| Ky: | Right here, this is going to be a little safe spot and they won’t know where I [meaning player] hide and so you [the player] can plan out what you are going to do next. And so when you are through, and you come out, they won’t be able to see you because there is a wall right here. So you can sneak around and stuff, but once you get here you have to go really fast because when they see you they are going to go after you. |

The above dialogue captured Ky’s intense interest on the user experience and the degree of agency he gave the third author. Most interesting perhaps is his inclusion of a safe space for the player to collect his or her thoughts. The creation of this element took forethought about potential users and how they may be feeling during game play. Ky established a dramatic stage on which the game-narrative played out, giving the third author the agency to rest and decide actions.

Additionally, Ky’s shifting of pronouns was perhaps interesting as well, reflecting the continual volley between second and third authorship he performed during coding third author agency into the game. While the lexical mistake might be considered random, the confusion might also be sourced in the dual roles the youth had to embody as they shifted back and forth from player to programmer. The player turns from “he” to “I” to “you” within the above dialogue. Such role acrobatics are not trivial and signal the complexity of authoring.

Connecting to the ever shifting nature of composing from multiple perspectives, all the youth had to also balance a “them” in the development of character and participant agency. *Them* refers to the degree to which allies and enemies were programmed to confine or assist the player’s action. Interestingly, some of the youth moved from being particularly unforgiving in the virility of their antagonists to developing allies to assist the player in the world. Zeb, for instance, went from making a game in which the player was blown away within the first few seconds of play to distributing the antagonists along a linear path that supplied the player with a more refined experience and actual agency to survive within the world for an extended period of time. Ky also introduced a character, the last surviving flying fish, to assist the user during play by attacking the flying fish that had turned into zombies. And finally, Ahmed created a deep structure of alliances and enemies that afforded help and constraint on play. He strategically positioned them on the terrain and in the air to make the play “fair for both side.” He spoke of the characters as if they were not attached to a player, though he considered how the player might be alerted to behaviors of alliance characters.

| | |
|---|---|
| Ahmed: | I am going to add more characters to this side of the field and program this blimp thing |
| Researcher: | What are you going to program it to do?<br>[Ahmed starts programming the balloon] |
| Ahmed: | When see cycles they glow to like warn the Kodus and the jet that they [the cycles] are coming to protect them. |

Although these illustrations are partly attributed to building an interesting story for the third author, they are also related to giving the characters and objects in the field of play agency either by adjusting behaviors or establishing relationships in opposition or in alliance. As Wardrip-Fruin (2009) suggested, “Designing experiences toward the satisfactions of agency involves balancing the dramatic probabilities of the world with the action it supports. In other words, the design task is to entice players to desires the game can satisfy” (p. 13). In this case, Ahmed was anticipating his players’ needs within the drama of the game. Satisfying the desire of an unknown player is a complex task in game making, where the interactions between the characters, the characters and landscape, and characters and user all need to be taken into account. The shifting of perspectives the youth underwent in the development of their games was complex. Considering character and participant agency was a key component in the making of a plot that was driven by character behaviors and landscape constraints and affordances, and not traditional plot lines. Since this compositional task was so complex, one is left to wonder how these strategies related (or not) with traditional academic domains, particularly writing. How is authorship with an Xbox controller and Kodu different from a pencil and paper or keyboard and screen? Despite obvious differences, there may be conceptual, complementary compositional underpinnings.

## Discussion and Educational Implications

Many educational researchers have presented descriptions on youth who are disengaged with learning traditional language arts curriculum but motivated by informal new literacy practices like videogames (Goodman, 2003; McGinnis, 2007; Ranker, 2008; Sarroub, Pernicek & Sweeney, 2007). Outside these depictions of the motivating and identity forming effects of New Literacies, however, there is little data on how engagement with various media domains intersects with what is defined as academic literacy.

Often times, traditional and New Literacies are considered "dueling discourses" (Alvermann, 2008), but in truth, while digital compositional spaces are unique, they also intersect with the goals of traditional writing instruction. Consider how the findings of Camp Kodu are connected to traditional literacy. Students crafted game-worlds in which they scripted character and player behaviors, motivations, weaknesses and strengths. They also produced settings that positively and negatively influenced character and player actions, and plot arcs were highly influenced by how the characters, players and settings were coded and designed. Additionally, the youth engaged and seemingly learned from the expectations associated with a genre and narrative plots, sometimes transforming both in interesting ways, much in the same vein as literary canons are built over time. They were also cognizant of audience expectations and afforded their users with agency through various controls of movement and expression. The students also demonstrated the importance of setting not just for mood and tone, but also as a space for user constraint and assistance. They developed characters that were interconnected and influenced each other's behaviors, and they continually inhabited the roles of author, users, and characters in order to create game-worlds that were appealing to an audience. Finally, the students constantly revised game text according to perceived audience wants and skill levels.

Centrality of audience, the roll of revision, and positioning one's authorship were essential to the students' creations and were motivated by the very academic goal of gaining and sustaining the attention of readers and viewers over time. During this authorship, the complex ways youth process information was largely implicit in the activity of game making. This processing overlaps with various literacy domains perhaps signaling a conceptual move away from learning as an explicit collection of knowledge to be acquired for mastery and a movement toward a framework based on the connections between knowledge, skills, and practices across authored spaced. Such a shift operates, according to Knobel and Lankshear (2008), on the vast and broad plane of encoding, not "letteracy," as literacy is often played out in schools, and it presses on idea of transliteracy (Thomas, Joseph, Laccetti, Mason, Mills & Perril, 2007).

Interestingly, learning transfer, as outlined by Schwartz, Bransford, & Sears (2005) suggested that multiple contexts and exposures to tasks underlie these implicit learning concepts. Extending their theories and findings to New Literacies and to programs like Kodu, it might be argued that compositional practices and skills are enriched by authoring in multiple modes of communication—generating an adaptive, compositional expertise that operates to some degree successfully across domains. For instance, by learning how composition operates across modes of communication, youth may build upon, learn new, and leave behind composing strategies as they discern what compositionally works from what does not across domains.

Consequently, future research in videogame making with Kodu might further unpack the relationship between authoring videogames and student reading and writing in traditional areas, looking at what transfers in and out of the practices associated these generative/creative activities across various modes and genres. Cross domain studies are needed in literacy education, and they suggest that videogame making through constructionist tools have merit outside of just teaching youth early programming concepts. As this small study itself demonstrated, students enlisted compositional strategies that were unique to game making, but also broadly related to authoring across compositional areas. Of course, dynamics of agency with Kodu seem uniquely situated in game making, though perhaps not without import to other compositional practices.

## References


Alvermann, D. (2008). Why bother theorizing adolescents' online literacies for classroom practice and research? *Journal of Adolescent and Adult Literacy, 52*, 8-19.

Barton, D. & Hamilton, M. (1994) Literacy practices. In D. Barton, M. Hamilton, & Ivonic, R. (eds.) *Situated Literacies: Reading and Writing in Context.* New York, New York: Routledge.

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Fribaum.

Burn, A. (2008). The case of rebellion: Researching multimodal texts. In J. Coiro, M., Knoble, C., Lankshear & D. Leu (Eds.) *Handbook of Research on New Literacy Studies* (151-178). New York, NY: Lawrence Earlbaum.

Cooper, S., Dann, W.,& Pausch,R. (2003). *Teaching objects-first in introductory computer science.* SIGCSE. Retreived September 17, 2009, from http://www.alice.org/publications/TeachingObjects-firstInIntroductoryComputerScience.pdf.

diSessa, A. A. and Abelson, H. (1986). Boxer: A reconstructible computational medium. Invited paper for a special issue on Teaching Programming, Communications of the ACM, 29(9), 859-868.

Fitzgerald, J. (1987). *Review of Educational Research*, Vol. 57, 4, 481-506.

Flower, L., Hayes, J.R., Carey, L. J., Shriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College, Composition, and Communication, 37.* 16-55.

Gee, J. P. (2003). *What videogames have to teach us about learning and literacy.* New York: Palgrave Macmillan.

Gee, J.P. (2007). Pleasure, learning, video games, and life: The projective stance. In M. Knobel & C. Lankshear (Eds.) A *New Literacy Sampler* (95-114). New York, NY: Peter Lang

Goodman, S. (2003). *Teaching youth media: A critical guide to literacy, video production, & social change.* New York: Teachers College Press.

Habgood, M., Ainsworth, S. & Benford, S. (2005). Endogenous fantasy and learning in digital games. *Simulation & Gaming, 36,* 483-498.

Hammer, J. (2007). Agency and authority in role-playing "texts". In M. Knobel & C. Lankshear (Eds.) A *New Literacy Sampler* (67-93). New York, NY: Peter Lang.

Hayes, E. (2008). Game content creation and its proficiency: An exploratory study. *Computers & Education,* 51, 1, 97-108.

Howland, K.L., Good, J. & Robertson, J. (2006). Script cards: A visual programming language for games authoring by young people. *Visual Languages and Human-Centric Computing* (VL/HCC'06), 181-186.

Ito, M., Horst, H., Bittanti, M, boyd, d., Her-stephenson, B., Lange, P., Pascoe, C.J., Robinson, L. (2008). *Living and learning with new media: Summary of findings from the digital youth project.* Retrieved Nov. 24, 2008 from http://digitalyouth.ischool.berkeley.edu/files/report/digitalyouth-WhitePaper.pdf.

Kafai, Y. B. (1993). *Minds in play: Computer game design as a context for children's learning.* Unpublished Doctoral dissertation, Harvard University.

Kafai, Y.B., Peppler, K.A., & Chiu, g. (2007). High tech programmers in low income communities: Creating a computer culture in a community technology center. In C. Steinfild, B. Pentland, M. Ackermann & N. Contractor (Eds.) *Proceedings of the Third International Conference on Communities and Technology* (pp. 545-562). New York: Springer.

Kelleher, C. & Pausch, R. (2006). Storytelling Alice motivates middle school girls to learn computer programming. Retrieved March 3, 2008 from http://portal.acm.org/citation.cfm?id=1240844.

Kenney, K. (2002). Building visual communication theory by borrowing from rhetoric. *Journal of Visual Literacy, 22*, 53-80.

Kodu. (2008). [Academic Version]. Redmond, WA: Microsoft Research.

Lankshear, C. & Knobel, M. (2000). *Strategies, tactics, and the politics of literacy: Genre and classroom practices in context of change.* Paper presented at the Third National Conference on Academic Texts. (Puebla, Mexico, April 15, 2000).

Lankshear, C. & Knobel, M. (2003). *New literacies, changing knowledge and the classroom.* London: Open University Press.

Knobel, M. & Lankshear, C. (2007). Sampling "the New" in New Literacies. In M. Knobel & C. Lankshear (Eds.) *A New Literacies Sampler* (pp. 1-24). New York: Peter Lang.

Knobel, M. & Lankshear, C. (2008). Remix: The art and craft of endless hybridization. *Journal of Adolescent & Adult Literacy, 52*, 22-33.

Left 4 Dead (2008). [XBox software]. Bellevue, WA: Valve Corporation.

Mackey, M. (2006). Digital games and the narrative gap. In D. Buckingham & R. Willett (Eds.) *Digital Generations: Children, Young People, and New Media.* Mahwah, NJ: Earlbaum.

McCutchen, D. (1994). The magical number three, plus or minus two: Working memory in writing. In J. S. Carlson (Series Ed.) & E. C. Butterfield (Vol. Ed.), *Advances in cognition and educational practice, Vol. 2. Children's writing. Toward a process theory of the development of skilled writing.* Greenwich, CT: JAI.

McCutchen, D., Francis, M., & Kerr, S. (1997). Revising for meaning: Effects of knowledge and strategy, *Journal of Educational Psychology, 89*, 667-676.

McGinnis, T. (2007). Khmer rap boys, X-men, Asia's fruits, and Dragonball Z: Creating multilingual and multimodal classroom contexts. *Journal of Adolescent & Adult Literacy, 50*, 570-579.

McKay, S.; Thurlow, C. & Zimmerman, H. (2005) Wired whizzes or techno-slaves? Young popel and their emergent communication technologies. In A. Williams & C. Thurlow (Eds.) *Talking Adolescence*. New York: Peter Lang.
Mateas, M. (2001) A preliminary poetics for interactive drama and games. *Digital Creativity*, 12.3, 140-152.
Miller, N. (2005) Wikipedia and the disappearing "author". *Etc.: A Review of General Semantics, 62*, 37-40.
Moskal, B., Lurie, D. & Cooper, S. (2000). Evaluating the effectiveness of a new instructional approach. Retrieved September 17, 2009 from http://www.alice.org/publications/EvaluatingTheEffectivenessOfANewApproach.pdf.
Nickerson, R. (1999). Enhancing creativity. In R. Sternberg (Ed.) *Handbook of Creativity*. Cambridge, UK: Cambridge University Press. 392-430.
Papert, S. (1969). See http://el.media.mit.edu/logo-foundation/.
Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
Pelletier, C. & Oliver, M. (2006). Learning to play in digital games. *Learning, Media, and Technology, 31*, 4, 329-342.
Peppler, K., & Kafai, Y. (2007a). What videogame making can teach us about literacy and learning: Alternative pathways into participatory culture. Retrieved March 3, 2008 from http://scratch.mit.edu/files/DiGRA07_games_kafai.pdf.
Peppler, K. & Kafai, Y. B. (2007b). From Supergoo to Scratch: Exploring creative digital media production in informal learning. *Learning, Media & Technology*, 32, 2. 149–166.
Peppler, K. & Kafai, Y. B. (2007c). Collaboration, computation, and creativity: Media arts practices in urban youth culture. In C. Hmelo- Silver & A. O'Donnell (Eds.), *Proceedings of the Conference on Computer Supported Collaborative Learning*, New Brunswick, NJ.
Perciles, K. (2007). *Game2Learn*. Retrieved March 3, 2008 from http://kpercicles.edublog.org/.
Ranker, J. (2008). Making meaning on the screen: Digital video production about the Dominican Republic. *Journal of Adolescent & Adult Literacy, 51*, 410-422.
Resnick, M., Kafai, Y., & Maeda, J. (2003). A networked, media rich programming environment to enhance technological fluency at after school centers in economically- disadvantaged communities. Retrieved March 3, 2008 from http://web.media.mit.edu/~mres/papers/scratch-proposal.pdf.
Resnick, M. & Silverman, B. (2005). Some reflections on designing construction kits for kids: Proceedings of Interaction Design and Children Conference, Boulder, CO.
Resnick, M., Kafai, Y., Maeda, J. (2003). A networked, media-rich programming environment to enhance technological fluency at after-school centers in economically-disadvantaged communities. Proposal to National Science Foundation.
Resnick, M., Maloney, J., & Rusk, N. (2006). Scratch and technological fluency. Retrieved August 12, 2009 from http://info.scratch.mit.edu/Research.
Robertson, J. & Howells, C. (2008). Computer game design: Opportunities for successful learning. *Computers & Education, 50*, 559-578.
Rusk, N., Resnick, M., Berg, R., & Pezalla-Granland, M. (2008). New Pathways into Robotics: Strategies for Broadening Participation. Retrieved Aug. 26, 2009 from http://web.media.mit.edu/~mres/papers/NewPathwaysRoboticsLLK.pdf.
Sarroub, L., Pernicek, T., & Sweeney, T. (2007). "I was bitten by a scorpion": Reading in and out of school in a refugee's life. *Journal of Adolescent & Adult Literacy, 50*, 668-679.
Schwartz, D., Bransford, J., & Sears, D. (2005). Efficiency and Innovation in Transfer. In J. Mestre (ed.) *Transfer of Learning*. Greenwich CT, IAP.
Thomas, S., Joseph, C., Laccetti, J., Mason, B., Mills, S., Perril, S., et al. (2007). Transliteracy: Crossing divides. *First Monday, 12*(12). Retrieved Sept. 16, 2009, from www.uicedu/htbin/cgiwrap/bin/ojs/index php/fm/article view 2060 1908.
Wardrip-Fruin, N. (2009). Agency reconsidered, again. Expressive Intelligence Student. Retrieved September, 1, 2009 from http://eis-blog.ucsc.edu/2009/08/agency-reconsidered-again/.
Willis, G. (1999). *Cognitive interviewing*. Retrieved on-line at on May 1, 2007 from http://appliedresearch. cancer.gov/ areas/cognitive/interview.pdf.

## Acknowledgments

We would like to thank Joan Davis, PhC, from the LIFE Center at the University of Washington for her assistance and reflections on Camp Kodu, as well as Deb McCutchen, PhD, for her expertise in research and writing.

# A Closer Look at the Split Attention Effect: Integrated Presentation Formats for Troubleshooting Tasks

Markus Huff, Vera Bauhoff, Stephan Schwan,
Knowledge Media Research Center Tübingen, Konrad-Adenauer-Str. 40, 72072 Tübingen
m.huff@iwm-kmrc.de, v.bauhoff@iwm-kmrc.de, s.schwan@iwm-kmrc.de

**Abstract:** According to the spatial contiguity principle of multimedia, learning interrelated information should be presented in an integrated instead of in a disconnected way. In this experimental study (80 participants), we examined performance with a troubleshooting task preventing such an integrated design. Instead, we used a display technology enabling a so-called "vexing image presentation", where there is no need for two foci of visual attention. Compared to a classic split screen setup in which participants have to shift their visual focus frequently, troubleshooting performance was superior in the "vexing image" mode. Hence, not the separation of external representations but processes of re-orientation after switching the visual focus seem to be responsible for split attention effects.

## Theoretical Background

As a well-established principle in multimedia design, the "spatial contiguity principle" is concerned with the spatial arrangement of different sources of information, for example a picture and an accompanying text description. Basically, the principle assumes that inter-related information (e.g., a geometric sketch and its related mathematical formula; Tarmizi & Sweller, 1988) should be presented in an integrated format in order to help the learners to mentally connect corresponding elements of the content (Ginns, 2006; Mayer, 1997). In contrast, if inter-related information is presented in a disconnected way (at locally distinct locations), learners have to search for corresponding pieces of information. This process wastes cognitive resources that are then missing for processes of elaboration and transfer into long-term memory (Mayer & Moreno, 2002). Another relevant example of the split attention effect is troubleshooting. Here learners have to find an error in a mechanical device and – if there is one – judge if this is critical for the proper functioning of the system. This is usually done by visually comparing a real device (or its depiction) with a reference device, be it another real device or a picture in a manual. Again, one finds the typical characteristics of split attention, namely, the necessity to switch the focus of attention frequently between two representations in order to mentally connect its corresponding elements in a time-consuming and error-prone manner.

However, although for combinations of text and pictures integrated presentations that avoid split attention are abundant (e.g., Scheiter, Gerjets, & Schuh, in press; Schnotz, 2002), until recently, similar solutions for combinations of real devices and pictures, or pairs of pictures, have been scarce. Accordingly, there have been few empirical studies to date that have examined split attention effects for learning tasks which require the comparison and mental integration of two pictures (Lee, Plass, & Homer, 2006; Huff & Schwan, subm.). But current digital technologies, such as augmented reality or auto-stereoscopic displays, offer new design possibilities for the problem of split attention. In particular, auto-stereoscopic displays allow for the realization of vexing images, that is, the presentation of two pictures in a way that enables participants to switch between them just by moving their head without a need for saccades (Huff & Schwan, subm.). Hence, one major goal of the present study was to examine the suitability of digital vexing images to substantially reduce the split attention effect for troubleshooting tasks that require extensive comparison of pairs of pictures.

Further, if one takes a closer look at the troubleshooting process, one can decompose it into at least two (distinct) sub-tasks, namely, a stage that can be termed as "comparative visual search" (Hardiess, Gillner, & Mallot, 2008) that is followed by a second stage which can be termed as "decision by mental animation".

During the first stage of troubleshooting, the learners have to identify one or more elements in the depiction of the target device that deviate from the respective parts of the correct reference depiction. In order to do this, correspondences between the elements of both pictures have to be established and for each corresponding pair, its similarity or dissimilarity has to be determined. This requires extensive visual search processes that are accompanied by frequent eye saccades and therewith shifts of the visual focus (Hardiess et al., 2008). Saccading alters the projection on the retina and consequently disrupts visual input. This in turn leads to a disruption of the visual working memory because the so-called "transsaccadic memory" store has a limited capacity and maintains visual characteristics of the learning material for only a brief time after stimulus offset (Phillips, 1974). If it would be possible to switch between different external representations without saccades, knowledge acquisition strategies could be based on the iconic memory store of high capacity (Irwin, 1991, 1992; Philips, 1974). In the present study, it is assumed that presenting the picture pairs as a vexing image via an auto-stereoscopic display might substantially reduce the need for shifts of visual focus and therewith the frequency and amplitude of saccades that are necessary to process the relevant information.

Once one or more possible problematic elements have been identified, the learner has to decide for each element whether or not it causes a malfunctioning. During this second stage, a promising strategy for diagnosing a breakdown could be to consider the identified element as part of a causal chain of events, and to simulate this chain of events by mental animation (Hegarty, Kriz, & Cate, 2003). Typically, mental animation is a process that poses high demands on cognitive resources. By reducing cognitive load during the first stage, more cognitive resources should be available for the second, decisional stage. Additionally, if participants can switch between two pictorial representations without moving their visual focus, the causal chain of events should also be easier to trace. Therefore, the second stage of processing should profit from an integrated presentation via vexing image, too.

Finally, an additional consideration concerns the kind of differences between a target device and its correct reference depiction. Here a distinction can be made between differences that are critical for the proper functioning of the target device (e.g., a missing gear in a clockwork) and differences that are not critical (e.g., a differentially colored part). Compared to functioning devices, the causal chain of events should be distorted and shorter for non-functioning devices. Therefore, identifying malfunctioning devices should be more accurate and faster than identifying a different but functioning clockwork. Hence, in sum, a second major goal of the study was to develop a more differentiated account of the cognitive processes that contribute to the split-attention effect.

### Hypotheses and Design

In the present study, participants were asked to solve a troubleshooting task by comparing two depictions of a mechanical pendulum clock and deciding whether one of the clocks was malfunctioning. After a learning phase in which the participants learned the basic principles of mechanical clockworks, picture pairs were presented either side by side in a classic split screen presentation mode (Split) or via an auto-stereoscopic display as a "vexing-image" in an integrated mode of presentation (Vex). While one of the pictures showed a functioning pendulum clock as the correct reference device, the second picture showed either the same (functioning) clock, or a different, but also functioning clock, or a different malfunctioning clock. The task was two-fold: First, learners had to visually search the picture pairs for differences and to state whether the two clocks were identical. If the two clocks differed, a second phase followed in which the learners had to decide whether or not the target clock was a functioning one.

We expected troubleshooting performance to be superior in the Vex condition for both the search and the decision sub-task. More specifically, visual search should be more accurate and faster in the Vex mode as there is no need to make a saccade when switching between the two external representations. This in turn should free some resources for the subsequent decision phase, again providing an advantage for the Vex mode. Further, we expected task performance to be superior for malfunctioning target clocks in the decision task, because in this case the causal chain of events is disrupted, which in turn should lead to faster and more accurate decisions.

Compared to the Split mode, where the two external representations were presented on two displays standing next to each other, in the Vex mode, the representations were presented on an auto-stereoscopic display. With such a display, participants could easily switch between two representations without shifting their visual focus. Instead, they just had to move their head. First participants, who were trained on the physical characteristics of a mechanical pendulum clock, had to accomplish a troubleshooting task. Second, they were asked to judge whether both clocks function properly. For this to work, we constructed three kinds of clockwork sets, namely, pairs of identical clockworks, and functioning and nonfunctioning pairs of non-identical clockworks. In the visual search sub-task, participants just needed to search for the respective identical and non-identical clockworks.

## Method

### Participants and Design

80 students of the University of Tübingen, Germany participated in the study. They received compensation for their participation and were randomly assigned to one of two cells of the between-subjects factor "presentation mode", which described whether information was presented in the Vex or in the Split mode.

### Materials and Apparatus

For each participant, the materials consisted of a computer-based learning environment that also included the troubleshooting tasks and a set of questionnaires, including one questionnaire for demographic information and a post-test containing transfer questions. Further we measured self-assessed cognitive load with a German adaptation of the NASA-TLX (Hart & Staveland, 1988), participants´ spatial abilities with the Mental Rotation

Test (MRT, Vandenberg & Kuse, 1978), and the Spatial Orientation Task (Hegarty & Waller, 2004). Visual span was measured with the Visual Patterns Test (VPT, Della Sala, Gray, Baddeley, & Wilson, 1997) and verbal span was assessed via the "Zahlen Merken" (digit span) scale from the Wilde Intelligence Test (Jäger & Althoff, 1983).

The computer-based learning environment was three-tiled. It took approximately 50 minutes to complete. First, participants were asked to learn the basic physical principles of a mechanic pendulum clock that were displayed on 7 HTML-pages using static images and short descriptive texts. More specifically, participants had to learn the functionalities of the weight as energy source, the escapement as central time base, and the different gears with the hands for indicating the time. The second part of this learning environment was a comprehension test that assessed participant's knowledge of pendulum clocks with 5 multiple-choice questions. For each question, three answers were displayed of which one was correct. (e.g., "Which statement is true?" (a) The ticking sound of a mechanical pendulum clock is generated by the pendulum. (b) The escapement is causal for the movement of the gears. (c) The weight is the energy source of the mechanical pendulum clock. Finally, the third part of this learning environment was the troubleshooting task. Participants were presented with two stills of a mechanic pendulum clock that were created using blender3D (http://www.blender.org). The two pictures either depicted exactly the same or different clockworks (see Figure 1). The "same" pictures showed clockworks that were exactly the same (including colors and the indicated time). This could be either two functioning clockworks or two malfunctioning clockworks. There were two kinds of "different" picture pairs. The functioning pairs showed different clockworks that both function properly but where one surface feature was different (e.g., the clocks indicated different times). In contrast, the malfunctioning pairs showed one functioning clockwork and one malfunctioning clockwork (e.g., a clockwork with a missing gear). Additionally, we varied the color of the clockworks presented pairwise across participants. Altogether, the troubleshooting task consisted of 40 pairwise clockworks (20 identical and 20 non-identical clockworks pairs). A training phase at the beginning of the troubleshooting task consisted of 8 clockwork pairs. These trials were excluded from further analysis.

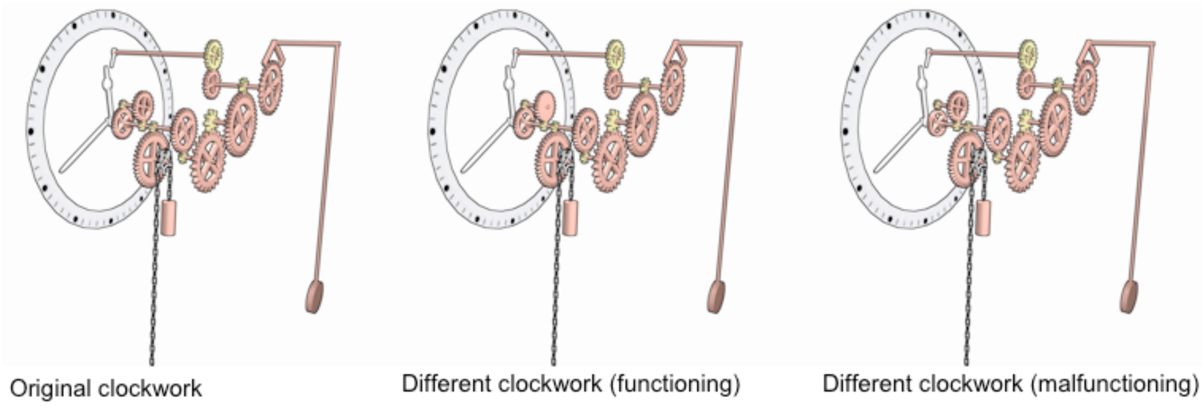



Figure 1. Examples of the stimulus material. In each trial, participants were presented with a combination of an original clockwork with a different clockwork (either functioning with a filled gear or malfunctioning with a gear without cogs).

There were two presentation modes: In the Split mode, the pictures were presented side by side on two displays standing next to each other. In contrast, in the Vex mode we used an auto-stereoscopic display that presented the two pictures such that participant were able to alternate between them simply by shifting their head. In this condition it is not necessary to shift the visual focus (see Figure 2; Huff & Schwan, subm.).

The experimental procedures were controlled by a PC and programmed using custom software programmed with python (http://www.python.org/). In both presentation modes, the first part of the learning environment was presented on a 24" display. The third part of the learning environment (troubleshooting task) was either presented on two 24" displays that were placed next to each other (Split mode) or on a single auto-stereoscopic display (Vex mode). In all presentation modes, the visual angles of the stills were comparable.

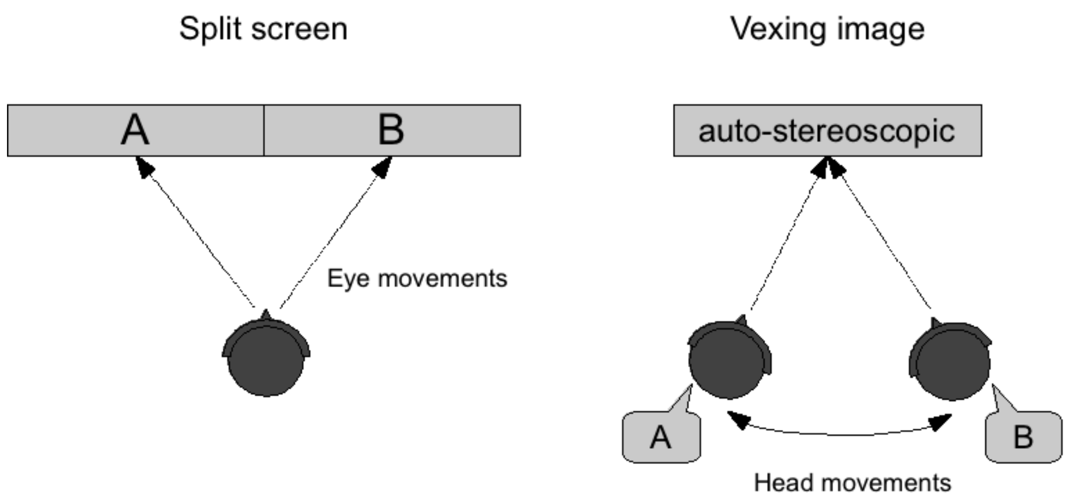



Figure 2. Split screen (left) and vexing image (right) presentation modes both presenting two pictorial representations ("A" and "B"). Whereas participants had to shift their visual focus in the Split mode, participants in the Vex mode just needed to move their head in order to see the second representation.

### Procedure

All participants were tested individually and received instructions via the computer monitor. At the beginning, participants completed the first two parts of the learning environment (learning phase and comprehension test). In the third part (troubleshooting task), participants were presented with 48 sets of clockwork pairs. Learners' task was two-fold. After the stills appeared on the screen, the participant had to visually compare the two clockworks and to answer the question whether both clockworks were identical by pressing the key "J" for "Yes" and "N" for "No, the clockworks are different.". Second, if the participant correctly detected that the two clockworks were different she/he had to decide whether both clockworks function properly or if one clockwork was defective. Again, the participant had to press the corresponding key in order to answer the question "Do both clockworks function properly?" ("J" for "Yes" and "N" for "No"). Half of the 40 experimental sets contained functioning and the other half non-functioning clockworks (see Table 1); the serial order and the left/right position of the original clockwork were chosen at random. The first 8 trials were for exercise purposes and were excluded from analysis.

Finally, the post-test, the mental-rotation test, the spatial orientation test, the visual patterns test, the test for the verbal memory span, and the cognitive load assessment questionnaire were filled out by the participants.

## Results

### Pre-Test

After the first part of the learning environment in which the participants worked through a description of the basic principles of clockworks, they received five multiple-choice questions about pendulum clocks in order to check their knowledge level. Each question had three possible responses of which one was true. Each correctly answered question was scored with one point. A total number of 5 points was possible. Three participants who obtained less than 3 points in this test were excluded from further analysis because they did not learn the relevant principles of a mechanical pendulum clock.

### Troubleshooting task

The troubleshooting task consisted of two components. In the comparison task, the participants had to decide whether the two depicted clockworks were identical or not. If the participants had detected a difference between the two clockworks, they were given a second task. In this task, they were required to decide whether the target clockwork was a functioning or a malfunctioning one.

#### Comparison task

Comparison task performance was measured with the sensitivity measure from the signal detection theory (Green & Swets, 1966) and search times to correctly identified sets of two original clockworks.
Mean hit rate (correct answers to target items – the clockworks were identical) and to false alarm rates (wrong answers to differing picture pairs – although the clockworks were different, participants answered "yes", both

for functioning and malfunctioning differing target clocks) were calculated for each participant in both conditions (Split vs. Vex). Next, two corresponding A' measures (Pollack & Norman, 1964) as sensitivity measure were calculated and submitted to a 2 (presentation mode; between-subjects) x 2 (target clock functionality; within-subjects) mixed factor ANOVA (see Table 1). Participants' performance was significantly influenced by both presentation mode and target clock functionality as the significant interaction of these factors showed, $F$ (1, 75) = 15.88, $p < .001$, $\eta_p^2 = .17$. A Bonferroni correction with $\alpha = 0.008$ was used to correct for the 6 multiple comparisons in this experiment. Performance in the "vexing image" mode was not affected by the kind of target clock type (functioning or malfunctioning, $p = .996$). In contrast, in the split screen mode, search performance was higher for malfunctioning target clocks than for functioning target clocks ($p < .001$). The main effect for "presentation mode" indicates higher performance in the Vex mode and the main effect for "target clock functionality" higher performance with malfunctioning clockworks, $F$ (1, 75) = 22.13, $p < .001$, $\eta_p^2 = .22$ and $F$ (1, 75) = 15.52, $p < .001$, $\eta_p^2 = .17$, respectively.

Table 1: Sensitivity measures of the comparison task (Question 1). Arithmetic means and standard deviations in parentheses.

| Presentation mode | Clock functionality | |
|---|---|---|
| | Functioning | Malfunctioning |
| Vexing image | .79 (.07) | .79 (.09) |
| Split screen | .69 (.06) | .74 (.05) |

The time from stimulus onset until the decision regarding the similarity of the clockworks was calculated for each participant in each condition (see Table 2). Data as presented in Table 2 were submitted to a mixed-factor ANOVA including the within-subject factor "clockwork type" (search times for identical, non-identical functional, and non-identical malfunctioning clockworks) and the between-subjects factor "presentation mode" (Vex, Split). Participants' search times were faster in the Vex mode as the main effect for presentation mode shows, $F$ (1, 68) = 38.03, $p < .001$, $\eta_p^2 = .36$. Further, clockwork type influenced search times too, $F$ (2, 136) = 6.14, $p = .003$, $\eta_p^2 = .08$. A Bonferroni correction with $\alpha = 0.024$ was used to correct for the 2 multiple comparisons in this experiment. Whereas there was no difference between the two non-identical clockwork types functioning ($M$ = 15.71 sec., collapsed over "presentation mode") and malfunctioning ($M$ = 15.95 sec., collapsed over "presentation mode"), $t < 1$, $p = .62$, search times for non-identical clockworks were faster than search times for identical clockworks ($M$ = 17.19 sec., collapsed over "presentation mode"), $t$s > 2.62, $p$s < .011. However, the interaction of clockwork type and presentation mode did not reach the level of significance, $F$ (2, 136) = 2.23, $p = .112$, $\eta_p^2 = .03$.

Table 2: Search times of the comparison task in seconds (Question 1). Arithmetic means and standard deviations in parentheses.

| Presentation mode | Clockwork type | | |
|---|---|---|---|
| | Identical | Functioning | Malfunctioning |
| Vexing image | 11.52 (6.08) | 10.83 (5.56) | 11.27 (7.79) |
| Split screen | 21.97 (7.01) | 19.81 (6.85) | 19.90 (6.55) |
| Mean (collapsed over "presentation mode") | 17.19 (8.40) | 15.71 (7.71) | 15.95 (8.31) |

## Decision task

If participants were presented with a pair of non-identical clockworks (with the target clock either functioning or malfunctioning) and they had correctly classified them as "not identical" in the comparison task, they were asked to state whether both clocks function properly. Again, we analyzed the data with the signal detection theory measure, sensitivity A', and reaction times for both correctly identified pairs of functioning clockworks and for correctly identified malfunctioning clockworks. Note, this task uses a two alternative forced choice paradigm. That is, the sensitivity measure as plotted in Table 3 reflects the participants' ability to distinguish between functioning and malfunctioning clockworks in the different presentation modes. Hence, it is not necessary to include this factor in the analysis.

Data as displayed in Table 3 were submitted to a t-test with the independent variable presentation mode for the sensitivity and to a mixed factor ANOVA with the between-subjects factor "presentation mode" and the within-subjects-factor “functionality” for the search times. Neither the sensitivity measure A', *t* (54.40) = 1.52, *p* = .134, nor the search times were influenced by the presentation mode *F*s < 2.65, *p*s > .11.

Taken together, both sensitivity and search time indicate that participants in the Vex mode outperformed participants in the Split mode. The superiority of the Vex condition is based on higher performances in the comparison task. In contrast, performance measures in the decision task were comparable across presentation modes.

Table 3: Results of the identification task (Question 2). Arithmetic means and standard deviations in parentheses.

| Presentation mode | Sensitivity (A') | Search time (seconds) | |
|---|---|---|---|
| | | Functioning clockworks | Malfunctioning clockworks |
| Vexing image | .78 (.07) | 4.99 (4.24) | 5.77 (7.82) |
| Split screen | .74 (.14) | 4.99 (4.34) | 6.64 (5.00) |

### Transfer test

In the final phase of the experiment, participants were asked to answer 6 multiple-choice questions that each described one malfunctioning clockwork (e.g., “The hour hand stands still, all other hands move correctly.”). Participants were presented with 3 alternative solutions (e.g., “The axis for the hour hand transmission is missing.”) of which one was true. Each correct answered question was coded with 1; maximum score was 6. Mean performance was 4.43 (*SD* = 1.14), there were no significant differences between presentation modes, *t* (74.90) < 1, *p* = .591.

### Cognitive load and learner related factors

Cognitive load was assessed with a German adaptation of the NASA-TLX (Hart & Staveland, 1988). The six questions were analyzed separately with a t-test including the independent variable presentation mode. Questions 1 (mental activity), 3 (time pressure), and 4 (effort) showed no significant differences between presentation modes (*t*s < 1.58, *p*s >.12). However, compared to the Split presentation mode, participants in the Vex mode reported a higher amount of physical activity (Question 2), *t*(64.45) = 2.51, *p* = .01, were more confident in scaling their own performance (Question 5), *t*(65.40) = 3.39, *p* = .001, and were less frustrated (Question 6), *t*(71.92) = -3.12, *p* = .003.

The experimental groups were checked for a-priori differences on the two learner-related factors spatial abilities (assessed with MRT and SO), visual memory span (VPT), and verbal memory span (assessed with ZM). We conducted a t-test for each of the tests separately with the independent variable presentation mode (Vex vs. Split presentation mode). There were no significant differences between groups with regard to spatial ability. Neither MRT, *t*(74.68) < 1, *p* = .996, nor VPT, *t* (74.99) = 1.31, *p* = .19, and SO, *t* (66.58) < 1, *p* = .874 reached level of significance. The same was true for working memory span, *t*(74.74) = 1.08, *p* = .28.

## Discussion

Although the spatial contiguity principle is well established in research on multimedia learning (Ginns, 2006; Mayer, 2001, 2006), few attempts have been made to extend the principle from picture-text combinations to picture-picture combinations. Hence, the present study attempted to develop an integrated format for the presentation of picture pairs in the context of a troubleshooting task and to compare it experimentally to a non-integrated split-screen presentation. Also, the troubleshooting task was decomposed into a two distinct phases, namely a “comparative visual search” phase (Hardiess et al., 2008), and a “decision by mental animation” phase.

We observed troubleshooting performance to be superior with a new display technology that allows for the integrated presentation of two pictures, compared to a traditional split-screen condition. The integrated presentation was realized by a vexing image which was presented via an auto-stereoscopic display. Here, it is not necessary to change the location of the visual focus in order to switch between the pictures. Instead, learners just had to move their head (see Figure 2). Hence, processes of reorientation were reduced to a minimum. In contrast to the split screen condition, visually comparing the two clockworks in the vexing image condition was both twice as fast and significantly more accurate. This finding suggests that error-prone processes of reorientation are one major reason for lower performance scores in learning environments with low spatial contiguity. In addition, participants were more confident in scaling their own performance and were less frustrated. If the split screen presentation mode can be characterized by frequent shifts of the visual focus and – as a consequence – frequent re-orientation processes, we can conclude that the split attention effect is reduced in the vexing image mode, which seems to be an adequate presentation mode for pairs of pictorial representations.

An interesting finding was that the functioning or malfunctioning of the non-identical clockworks did influence troubleshooting performance in the Split mode but not in the Vex mode. Because viewers in the vexing image mode could keep their visual focus at the critical position while switching between the representations, one can assume that the comparison task can be solved primarily by means of perceptual pattern matching. In contrast, in the Split mode participants cannot rely on perceptual information alone. Switching between representations requires also memory based processing. Participants have to encode parts of one representation and compare this mental representation with the corresponding part of the second representation. Such memory-based processes might involve mental animating of the static clockworks (Hegarty et al., 2003). A central part of mental animation as described by Hegarty et al. (2003) is the identification of the causal chain of events. Learners have to identify the energy source and the transmitting gears. This causal chain is disrupted in malfunctioning clockworks. As participants were instructed in the mechanical principles of pendulum clocks at the beginning of the experiment, this might have helped them during troubleshooting.

Surprisingly, we observed all these effects in the first phase when participants were asked to visually search for differences between the clockworks. Originally, we hypothesized that such processes take place in the decision phase and not – as observed in this study - during the visual search phase. One reason for this might be that participants intermixed the tasks of comparing the clocks and deciding whether the clocks work properly. If this is true, the decision about the proper functioning of the clockworks might have been finished before the decision task began and participants used the decision phase just to verify their decision. For a clearer distinction between those two phases, measures of process related variables such as gaze behavior could be employed. Further research should include eye gaze as dependent measure in order to gain deeper insights in the cognitive processing of separated information.

Finally, we measured participants transfer knowledge in a verbal-based transfer test. As there was no need for costly re-orientation processes in the Vex mode, we hypothesized that participants should be able to encode a more elaborate and more precise mental representation of the mechanical clockwork. As a consequence, compared to participants in the Split mode, transfer test performance should be higher. However, transfer test performance was comparable across the two presentation modes. Faster search times in the Vex mode indicate that learners did not use the free mental resources to encode an elaborated mental model. Instead, they used the free capacity to speed up their responses.

## Conclusions

In conclusion, we were able to show that the process of troubleshooting as employed in this study is highly dependant on visual search processes. Shifting the visual focus frequently involves costly processes of re-orientation. The subsequent decision, whether or not the identified difference between two mechanical devices is critical for its proper functioning, is independent of the presentation mode. Solving this task does not require any shifts of visual attention. Instead, learners can solve it by mentally animating both clockworks in a row. They are finished if they find an error or if they successfully animated both clockworks. An integrated design not only reduces the number of shifts of visual focus but also leads to lower frustration scores and higher confidence ratings. Hence, multimedia design reducing costly re-orientation processes is beneficial for the troubleshooting task as a whole. This finding has also consequences for our understanding of multimedia learning. Whereas previous research focused mainly on the combination of pictorial and verbal information, this study showed that common principles of multimedia learning also apply to purely pictorial material. Finally, this study also demonstrated that new display technology is helpful in designing new kinds of learning environments and, in turn, in gaining new insights in cognitive processes involved in learning from multiple external representations.

## References


Della Sala, S., Gray, C., Baddley, A.D., and Wilson, L. (1997). *Visual Pattern Test*. Bury St. Admunds: Thames Valley Test Company.

Ginns, P. (2006). Integrating information: Meta-analyses of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511-525.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford: Wiley.

Hardiess, G., Gillner, S., & Mallot, H.A. (2008). Head and eye movements and the role of memory limitations in a visual search paradigm. *Journal of Vision, 8,* 1-13.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Hg.), *Human mental workload*. (pp. 139-183). North-Holland.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction, 21*(4), 325-360.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32,* 175-191.

Huff, M., & Schwan, S. (subm.). Integrating information from two pictorial animations: Complexity and cognitive prerequisites influence performance.
Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*(3), 420-456.
Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*(2), 307-317.
Jäger, A. O., & Althoff, K. (1983). *Der WILDE-Intelligenztest (WIT)* [The WILDE intelligence test]. Göttingen, Germany: Hogrefe.
Lee, H., Plass, J. L. & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology, 98*(4), 902-913.
Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32,*1-19.
Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.
Mayer, R. E. (2006). *Cambridge Handbook of Multimedia Learning*. New York: Cambridge University Press.
Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review, 14,* 87–99.
Phillips, W. A. (1974) On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*, 283-290.
Pollack, L. & Norman, D. A. (1964). Non-parametric analysis of recognition experiments. *Psychonomic Science, 1,* 125–126.
Scheiter, K., Gerjets, P., & Schuh, J. (in press). The acquisition of problem-solving skills in Mathematics: How animations can aid understanding for structural problem features and solution procedure. *Instructional Science*.
Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review, 14*(1), 101-120.
Tarmizi, R. & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*, 424-436.
Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations, a group test of three dimensional spatial visualizations. *Perceptual and Motor Skills, 47*, 599-604.

## Acknowledgments

Markus Huff is indebted to the LANDESSTIFTUNG Baden-Württemberg for the financial support of this research project by the Eliteprogramme for Postdocs. The authors thank Kathrin Stalder and Marco Dehez for their help in conducting this experiment.

# Facilitating Group Learning in Science Laboratory Courses Using Handheld Devices

Chen-Wei Chung[a], Wang-Hsin Kuo[b], Chen-Chung Liu[b]
[a] *Department of Computer Science and Engineering, National Central University, Taiwan*
[b] *Graduate Institute of Network Learning Technology, National Central University, Taiwan*

**Abstract:** The laboratory course is an inferential curriculum designed to foster science inquiry ability in science education. However, in the universities of Taiwan, laboratory courses often last only a limited time and explore only a single topic of inquiry. Students' individual ideas and experiences are not well represented in the group inquiry process. Therefore, this study attempts to provide an organizing tool on PDAs to help student groups regulate their inquiry processes and engage in reflective inquiry activities.By gathering data from 56 students (17 groups) participating in the laboratory course, this study utilizes handheld devices as an observational tool for peer interactions in collaborative inquiry activities. A total of 360 minutes of activity log was analyzed for details of peer interaction in order to understand how handheld devices can influence group regulation when students are provided with group regulation functions on PDAs. Students demonstrated six main types of interaction when formulating group inquiry plans via PDAs. This study makes distinction between four patterns of cooperative PDA use. Such distinction could assist teachers and system designers in facilitating group activities by avoiding unfavorable interaction patterns. The results show that unperceived conflicts concealed in individual devices impede peer interaction. This finding suggests that designers of learning environments should consider the importance of reinforcement mechanisms applied in Mobile Computer Supported Collaborative Learning (MCSCL) to maintain shared understanding.

**Keywords:** Laboratory courses, group regulation, handheld devices, peer interaction analysis

## 1. Introduction

The laboratory course follows an inferential curriculum designed to foster science inquiry ability in science education (Hinrichsen & Jarrett, 1999; Hollingworth & McLoughlin, 2001). Students in laboratory courses learn how to pursue scientific knowledge through inquiry and experience the process of constructing such knowledge (Edelson, Gordin, & Pea, 1999). Within the inquiry process, students formulate testable hypotheses, measure variables, and design experiments. However, in current practice, the design and implementation of laboratory courses is based on an instructional rather than hands-on approach (Mayer 2004; Kirschner, Sweller , & Clark, 2006), i.e., students complete experiments step by step according to the textbook or an instructional manual provided by the teacher. Such an approach deprives students of the opportunity to experiencethe reflective inquiry process and to learn how to acquire knowledge through that inquiry process in laboratory courses and settings.

Collaborative inquiry has recently come to be considered an effective mode of reflective scientific inquiry. Students can actively develop problem solving and knowledge construction activities during peer discussion and interaction (Blumenfeld, Marx, Soloway, & Krajcik, 1996; Littleton & Hakkinen, 1999). Peer discussion and interaction can facilitate social construction of knowledge (Jonassen, 1994) that cannot easily be attained through individual inquiry learning. However, laboratory courses often last onlya limited amount of time and explore only a single inquiry topic. For example, laboratorycourses in the universities of Taiwan consist only of weekly 3-hour classes covering a single topic. Students' individual ideas and experiences are not well represented in the group inquiry process due to time limitations. Thus, students may fail to get the opportunity to exchange ideas and experiences with peers. Furthermore, previous studies have confirmed that students still have difficulty formulating testable hypotheses, designing conclusive experiments, and compiling compatible data (Manlove, Lazonder, & De Jong, 2006). In order to achieve effective reflective scientific inquiry, student groups need regulative support in collaboratively forming inquiry strategies. Therefore, it is necessary for laboratory courses to provide an organizing tool to help student groups regulate their inquiry processes and conduct reflective inquiry activities.

Prior to this study, researchers (Manlove, Lazonder, & De Jong, 2006) had developed some regulative learning tools on handhelds designed to help students with collaborative development of inquiry strategies; handhelds had become increasingly affordable in classrooms over the course of recent years (Roschelle, 2003; Zurita & Nussbaum, 2004; Liu et al., 2006; Fry & Hin, 2006; Liu & Kao, 2007), and increased student participation and collaboration in classrooms had been experimentally facilitated by means of handheld devices. However, although students could use handhelds as regulative tools in their pursuit of knowledge, they had difficulty generating a complete group experience in the limited time available. They could not easily exchange experiences with peers due to lack of individual inquiry experience. Therefore,this study designed an inquiry learning activity which integrated in-class discussion of inquiry planning using PDAs with pre-class regulation of inquiry planing. Students could organize individual exploration plans on the web to build individual inquiry experience in advance. At the laboratory, handhelds were adopted as a group regulation tool to help students organize and discuss their group inquiry plans based on their individual inquiry plans. This study attempts to bridge the gap between individual and collaborative construction of knowledge in order to promote efficiency and effectiveness in science laboratory courses.

This study attempts to understand how handheld devices can influence group regulation when students are provided with group regulation functions on PDAs. Recent research methodologies in Computer Supported Collaborative Learning (CSCL) have put significant emphasis on the micro analysis of peer interaction, through which improvement of learning may take place (Stahl, Koschmann, & Suthers, 2006). This study not only takes into consideration the performance gain resulting from technologies but also attempts to uncover the reason why the technology improves the interactions that lead to meaningful learning (Liu et al, in press). Accordingly, this study uses micro analysis to explore the interaction demonstrated by student groups using the group regulation functions of PDAs. Furthermore, to provide a comprehensive understanding of how the PDAs and regulation functions influence group inquiry achievement in laboratory courses, this study also analyzes the relationship between group inquiry achievement and interaction patterns as revealed by micro analysis. By gathering and analyzing the system usage logs of 56 college students in a laboratory course in National Central University, this study attempts to answer the following research questions based on peer interaction analysis:

- How do the students in a group interact with each other through PDAs with group regulation functions when using pre-regulated inquiry plans to form a group plan?
- What kind of peer interaction patterns can be observed when students develop regulative strategies using PDAs that have group regulation functions?
- How do the peer interaction patterns facilitated by the PDAs influence group inquiry achievement?

## 2. Group inquiry planning via handhelds

The laboratory course involved in this study included a pre-class inquiry regulation activity and an in-class collaborative inquiry regulation activity. The pre-class inquiry activity required students to build individual inquiry plans based on a given science topic. Students were thereby able to learn the procedure and objectives of an inquiry activity. In the in-class inquiry regulation activity, students were required to work with their partners to generate a group plan based on the pre-generated plans. This study developed a pre-class inquiry regulation tool on the web. Using this tool, students could make an inquiry plan. An inquiry plan consists of six main items: (1) purposes of the inquiry, (2) related theories of the inquiry, (3) instruments, (4) hypothesis, (5) measurement variables, and (6) procedures.

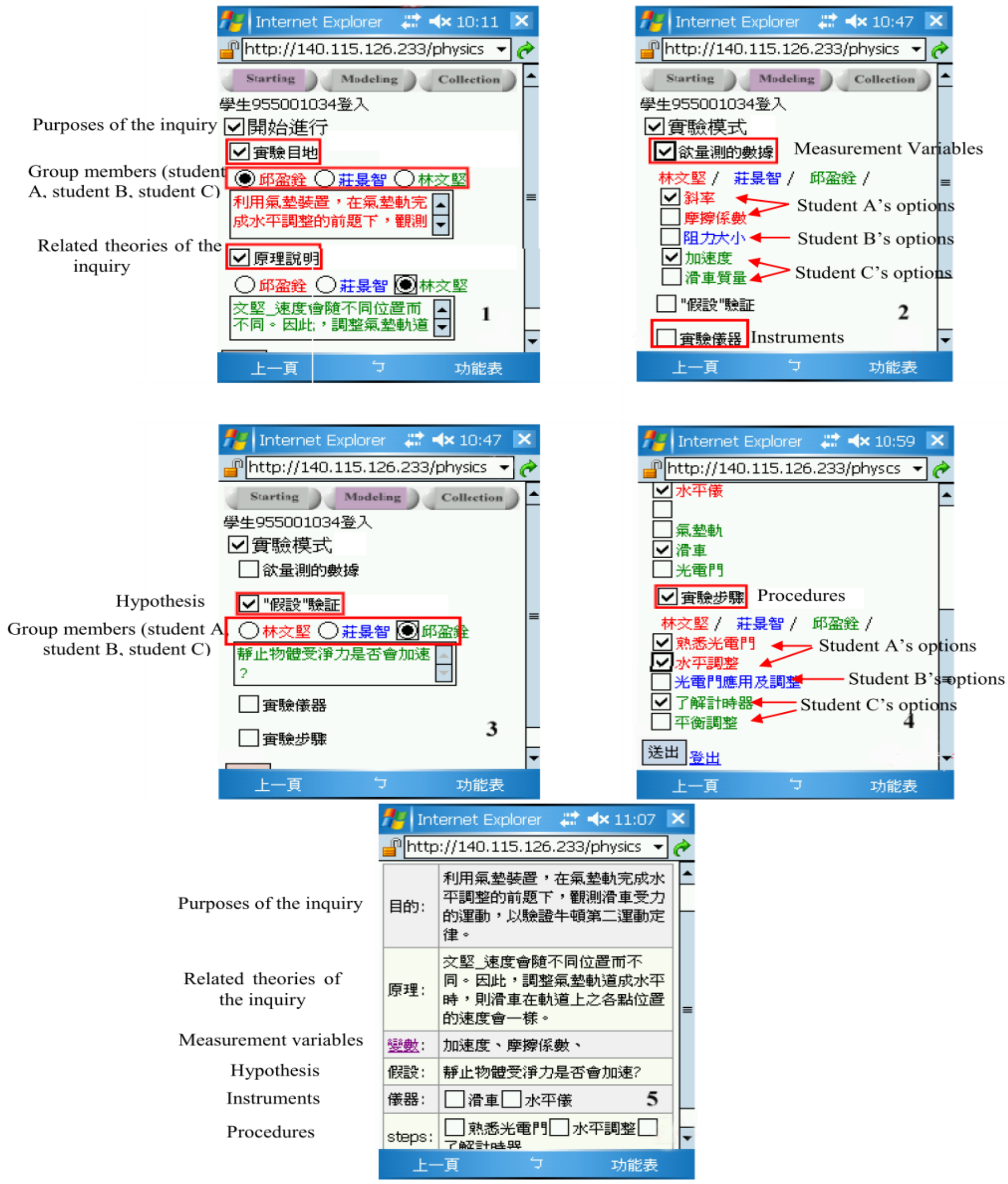



Figure 1. The regulative tool for group inquiry planning on PDAs

In the laboratory, the teacher would outline the purposes of inquiry, and then student groups would start to regulate their group inquiry plan, which would contain the same six main items found in the individual inquiry plans. In order to formulate testable hypotheses and design conclusive experiments, students were asked to collaboratively revise and discuss the individual inquiry plans they had generated on the course website to reach a unified group plan. The handhelds were used as an organizing tool to combine students' individual plans into a group inquiry plan. Using the PDAs, students could review their own inquiry plans and compare them with those developed by their partners while deciding on a unified group plan. An application on the PDAs could generate a summary of the group inquiry plan based on the decisions the students made in class. Students could then proceed to the inquiry activity and gather data based on the unified group plan. This whole process is interactive, because students were able to modify the group plan during the inquiry if necessary. After collecting all the data necessitated by the group inquiry plan, students would discuss and develop a final report according to the inquiry plan summary generated by the PDAs. Finally, the teacher would evaluate the inquiry achievement of the student groups based on the final reports.

This study develops a regulative tool for group inquiry planning on PDAs intended to support group members in comparing, reviewing, and generating unified inquiry plans (Fig. 1). The regulation tool enables students to use PDAs to refer to each others' pre-designed individual inquiry plans. During class activities, the individual inquiry plans students made before class became options that all students could choose from when selecting the six main items for their group inquiry plans. Students could review and check the provided options for the main items on the PDAs. In order to generate a summary of an inquiry, firstly students needed to decide on the purposes and related theories of the inquiry. Individuals' plans regarding the above main items were shown as options in different colors (Step 1 in fig. 1). Students could distinguish between and compare the different options provided by the group members by color. Secondly, students needed to decide which of the measurement variables proposed by group members should be included in the group inquiry plan Students were able to do this using the PDAs by checking boxes next to the variables (Step 2 in fig. 1). Thirdly, students needed to select a hypothesis for the inquiry from among the hypotheses proposed by the group members (Step 3 in fig. 1). Fourthly, students needed to decide on an executable procedure for the inquiry from the list of procedures proposed by the group members (Step 4 in fig. 1). After students had finished the group regulation activities, the tool would generate a summary of their inquiry plan on the PDAs (Step 5 in fig. 1). These interactions were recorded on the PDAs along with details of the peer interactions that took place during the group regulation and inquiry activity. Students' activities on the PDAs were archived for later analysis of peer interaction so that the effect of the PDAs' group regulation tool on learning experience in laboratory courses could be confirmed.

## 3. Method

The primary purpose of this study is to explore the effects of handheld devices with the above mentioned regulation functions by analyzing the peer interactions involved in collaborative inquiry activities. The participants were 56 freshman students enrolled in a physics laboratory course at National Central University (Taiwan). Students were divided into 17 groups, whose size ranged from three to four students. This study analyzes student interactions over the course of two 3-hour sessions of a weekly laboratory class, during which they performed experiments focused on Newton's second law of motion and the law of conservation of angular momentum. Students were able to make their inquiry plans on the course website provided by the study. During the laboratory class, students modified their inquiry plans, referred to options presented in other group members' inquiry plans, and regulated a unified group plan using the PDAs provided by this study. All of the students' interactions were recorded by the PDAs while the students regulated their group inquiry plan. After the class, the teacher evaluated the inquiry reports.

Figure 2 shows the collaborative inquiry behavior found in two threads as recorded by the PDAs. For instance, if student 1 (s1) checks item 1, this behavior is denoted as s1 (1). Likewise, if student 1 checks the second option of item 3, this is denoted as s1 (3_2). All regulation activities of individuals, such as modifying options, choosing options presented by other group members, and reviewing/comparing options provided by group members, were logged in the PDAs. The activity log helps demonstrate the students' regulation behaviors and development of regulative strategies as aided by PDA-based tools. This study analyzes peer interactions to determine what peer interaction patterns students can engage in through PDAs and to confirm the influence of the regulation functions provided by PDAs on the effectiveness of inquiry activities.

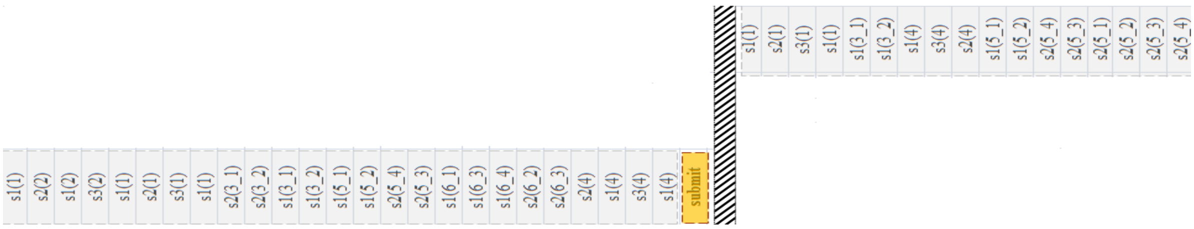



Figure 2. The activity log

As students performed collaborative regulation activities in the laboratory, data were collected and recorded by the PDAs. This study analyzes a total of 360 minutes of activity log. All analyses of peer interactions were performed by two independent researchers (coders). The analyses included the identification of peer interaction activities, e.g. comparison, conflict, and resolution, as well as the classification of peer interaction patterns. The inter-coder reliability (agreement) for each analysis was at least 85%, indicating that the analysis is sufficiently reliable. Researchers resolved disagreements in analysis and categorization through discussion.

## 4. Results and discussion

The results of the analysis are categorized to address the following questions. (1) What types of peer interaction can PDAs facilitate in order to help students to regulate their plans? (2) What general patterns of peer interactions can be observed in students using PDAs? (3) How do the general patterns of peer interaction affect student achievement in the context of the inquiry activities required by the laboratory course?

### Peer interactions through PDAs

The activity logs were analyzed in detail to explore the peer interactions that occur during group regulation activity. Upon analysis of the activity log, six main types of interaction were identified within the group regulation activities. These types are as follows:

- Comparison: Students often compared their own individual inquiry plans with those of others by switching the options of some items. In general, they performed such activities in a short period of time in order to understand the inquiry plans of others.
- Perceived conflict (change of options): Such interaction occurred when one student chose a different option from that which was chosen by another group member. This behavior indicates that the student has noticed a difference of opinion between himself/herself and others regarding the group inquiry plan. For instance, if student A chooses the hypothesis proposed by student B as the hypothesis of the group inquiry plan, student C might modify the group inquiry plan by choosing another hypothesis proposed by himself/herself, creating a perceived conflict between student A and student C.
- Perceived agreement: Such an interaction occurs when a student reviews the group plan and agrees on the group inquiry plan as proposed by others.
- Resolution of conflict: While regulating a group plan, student A might switch to the option that student B chose after detecting the conflict between B and himself/herself (perceived conflict) regarding that item in the plan. Students A and B would then have reached a resolution regarding this item.
- Unperceived conflict: The PDAs are personal devices that allow individuals to modify inquiry plans. There generally exists a time lag between personal modification of the application server and acknowledgment of the modification on other members' PDAs. For example, one student might modify the group inquiry plan with his/her PDA and thus cause a conflict with the group plan proposed by others. If other students cannot detect the conflict before this inquiry plan is sent off to others, this is unperceived conflict. Since the inquiry plans were concealed in the PDAs, other students sometimes could not detect the conflicts between themselves regarding the inquiry plan.

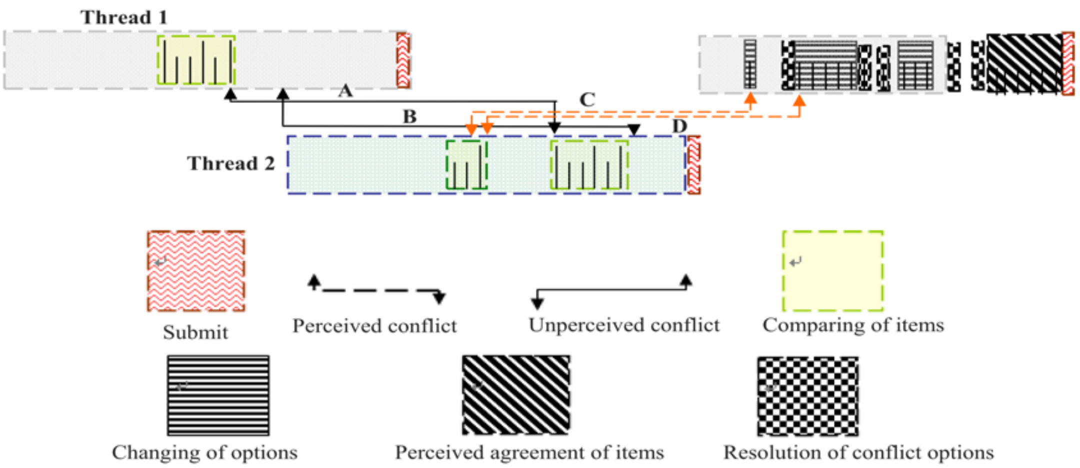



Figure 3. Group dynamic chart

In order to depict peer interaction explicitly, this study codifies and represents students' interactions based on the aforementioned peer interaction events in a group dynamic chart (Fig. 3). Each bar in a row represents a thread of student regulation activity in the PDA. Each dotted arrow from one thread to another in the chart denotes a perceived conflict. For example, Fig. 3 shows two threads that proceeded concurrently after the student in thread 2 submitted his plan. The other student in thread 1 checked a different option for an item and noticed the conflict between his inquiry plan and those of others (e.g.: line C and line D). Therefore, there is a dotted arrow between the two threads to represent the perceived conflict. In addition, each thick arrow between threads in the chart denotes an unperceived conflict. For instance, the arrow between threads 1 and 2 in the first chart of Fig. 3 reveals that a student had not noticed another student's inquiry plan and did not perceive the difference between that plan and his/her own (e.g.: line A and line B), because the student in thread 1 had not submitted his/her plan. A rectangle with vertical lines represents a student comparing his/her plan with those of others. Each long vertical line indicates a student checking his/her own plan, while a short line indicates him/her checking the plans of others. A rectangle with a horizontal line shadow denotes conflict and modification. A rectangle with a diagonal line shadow denotes conflict resolution. Perceived agreement is indicated by a rectangle with a dotted shadow. The group dynamic chart of each group was analyzed to determine what interaction patterns occurred during the collaborative inquiry activities.

Analysis of the group dynamic chart reflects the frequency of different peer interaction activities. Table 1 lists the frequency of each type of peer interaction. Table 1 reveals that PDAs helped students obtain a global view of each individual's inquiry plan. Students frequently performed comparison activities using the PDAs. In total, students performed such comparisons on 152 items in the different individual inquiry plans. In addition, the PDAs also helped students to express agreement with their group members. While comparing different options, students showed perceived agreement in 94 items of the group inquiry plans. In addition, students detected perceived conflicts in 75 items and chose 169 options which differed from those chosen by their peers. Of the changes of option selection in the group inquiry plans, 65 options were ultimately determined to be consistent among group members. Thus, there were still many conflicts that remained unresolved. Interestingly, unperceived conflicts occurred with high frequency. The log generated by the PDAs shows that there were conflicts between members over 79 items. This indicates that conflicts among students are often concealed when PDAs are used to mediate the formation of group plans. When using such a mechanism, students still have difficulties solving conflicts when attempting to create a unified group plan.

Table 1. Frequency of peer interaction in collaborative inquiry regulation activity

| Interaction types | Frequency |
|---|---|
| Comparing of items | 152 items |
| Perceived agreement | 94 item |
| Perceived conflict (changing of options) | 75 item (169 options) |
| Resolution of conflict | 65 (options) |
| Unperceived conflict | 79 items |

## Peer interaction patterns through PDAs

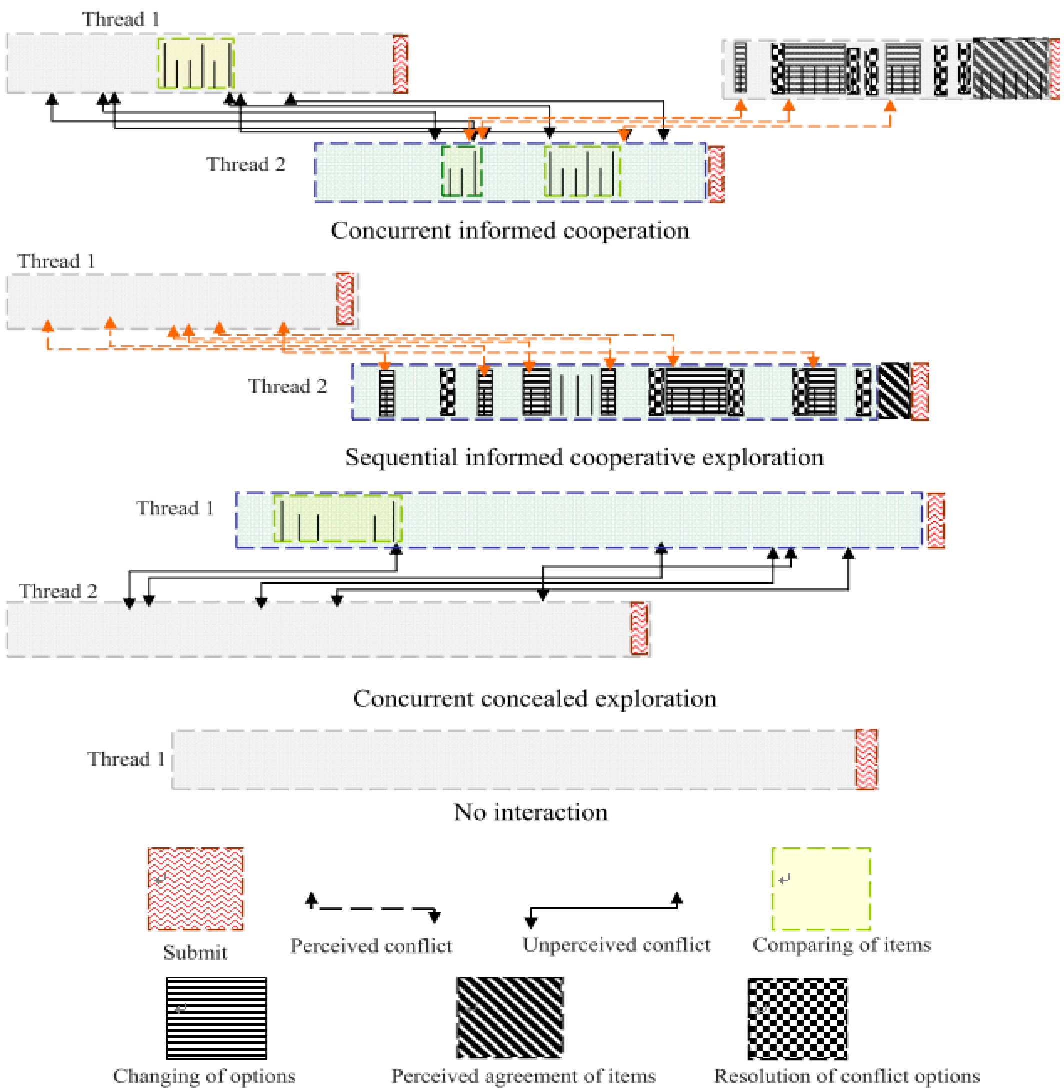



Figure 4. The group dynamic chart of four peer interaction patterns

Analysis of the group dynamic charts identifies four distinct interaction patterns (Fig. 4) occurring during the collaborative inquiry activities observed in this study. These patterns are as follows:

- *Concurrent informed cooperation*: This pattern takes the form of two concurrent threads and occurs when students begin by working independently and then jointly revise each other's plans to form the final plan. At the beginning of this pattern of interaction, individuals do not inform each other of the work they are doing, but instead each organizes a group plan individually. This behavior allows for frequent occurrences of unperceived conflicts at the beginning due to the concurrent nature of the participation. At the later stage of the pattern, students revise their group plans based on the individuals' results. Perceived conflicts are detected and resolved only at this stage. Therefore, the main characteristic of this pattern is the initially independent organization of group plans, upon which students then jointly base further revision of their group plan.

- *Sequential informed cooperation:* This interaction pattern occurs when students organize their plans sequentially, i.e., one student initiates a new thread and modifies the group plan, and then another student further revises the group plan based on these modifications. Unlike concurrent informed cooperation, this pattern only allows for perceived conflicts. There are no unperceived conflicts between the two threads.
- *Concurrent concealed cooperation:* Students following this pattern organize their plans concurrently and independently. They do not notice the difference between their own inquiry plans and those of others. The conflicts between their inquiry plans are concealed, and thus many unperceived conflicts occur in this pattern. Consequently, students who exhibited this pattern did not demonstrate any conflict resolution.
- *No interaction:* This pattern occurs when students do not interact with each other to make their group plans.

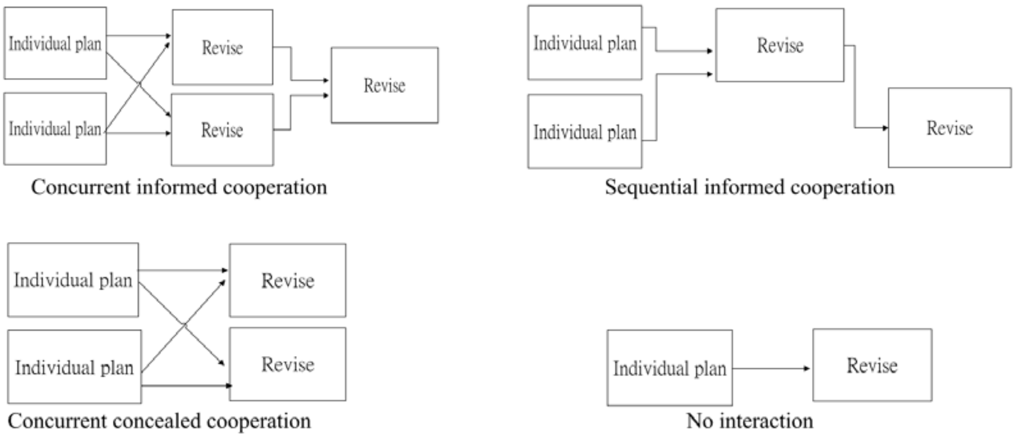



Figure 5. Peer cooperation patterns of the four interaction patterns

The four interaction patterns defined by this study demonstrate how students work together to organize their group plans using PDAs. Figure 5 presents the four peer cooperation patterns in a comprehensive manner. In concurrent informed cooperation, students generated group plans based on all individual inquiry plans and then revised the plan collaboratively. Students exhibiting sequential informed cooperation generated a group inquiry plan and then each revised the plan one after another. Less productively, using the concurrent concealed cooperation pattern, some students generated group plans based only on their individual plans without any cooperation. Even more disconcertingly, some students did not interact with each other at all when formatting their group plans, in accordance with the no interaction pattern.

This study categorizes the interaction patterns of each group based on the group dynamic charts. The distribution of interaction patterns is represented in Table 2. Only five groups demonstrated concurrent informed cooperation while thirteen groups exhibited the pattern of sequential informed cooperation. Surprisingly, there were ten groups working in the pattern of concurrent concealed cooperation and six groups working in the no interaction pattern; that is, students in these groups worked independently even when the regulation tools of the PDAs were provided to facilitate group regulation of plans.

Table 2. Group achievement distribution among different interaction patterns and interaction behaviors mediated by PDAs

| Group | Type of cooperation | Planning time | Score | Comparison of items | Change of options | Unperceived conflict | Perceived agreement of items | Resolution of conflict options | Resolution ratio |
|---|---|---|---|---|---|---|---|---|---|
| 6 | Concurrent informed cooperation | 00:22:43 | A+ | 5 | 7 | 4 | 6 | 3 | 43% |
| 1 | | 00:25:36 | A | 5 | 5 | 6 | 9 | 4 | 80% |
| 20 | | 00:21:49 | A | 12 | 9 | 6 | 6 | 5 | 56% |
| 2 | | 00:30:29 | A- | 6 | 4 | 5 | 10 | 3 | 75% |
| 18 | | 01:45:50 | A- | 7 | 10 | 6 | 6 | 5 | 50% |
| 23 | Sequential informed cooperation | 00:20:40 | A+ | 5 | 9 | 0 | 4 | 4 | 44% |
| 3 | | 00:21:23 | A- | 5 | 5 | 0 | 6 | 2 | 40% |
| 22 | | 00:22:40 | A- | 5 | 13 | 0 | 3 | 5 | 38% |
| 5 | | 00:23:52 | A- | 0 | 13 | 0 | 3 | 5 | 38% |
| 7 | | 00:15:31 | B+ | 0 | 8 | 0 | 8 | 3 | 38% |
| 29 | | 00:14:04 | B+ | 6 | 11 | 0 | 2 | 5 | 45% |
| 31 | | 00:13:26 | B+ | 0 | 15 | 0 | 1 | 7 | 47% |
| 11 | | 00:18:39 | B+ | 0 | 10 | 0 | 5 | 3 | 30% |
| 12 | | 00:16:28 | B+ | 5 | 11 | 0 | 6 | 3 | 27% |
| 14 | | 00:13:41 | B+ | 5 | 10 | 0 | 3 | 2 | 20% |
| 16 | | 00:16:32 | B+ | 0 | 9 | 0 | 7 | 0 | 0% |
| 19 | | 00:17:48 | B | 6 | 6 | 0 | 9 | 2 | 33% |
| 21 | | 00:12:17 | B | 5 | 14 | 0 | 0 | 5 | 36% |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4 | Concurrent concealed cooperation | 00:08:23 | B | 5 | 0 | 5 | 0 | 0 | 0% |
| 8 | | 00:09:26 | B | 4 | 0 | 5 | 0 | 0 | 0% |
| 9 | | 00:10:35 | B | 5 | 0 | 5 | 0 | 0 | 0% |
| 10 | | 00:12:46 | B | 5 | 0 | 6 | 0 | 0 | 0% |
| 13 | | 00:09:28 | B | 5 | 0 | 5 | 0 | 0 | 0% |
| 24 | | 00:09:01 | B | 6 | 0 | 5 | 0 | 0 | 0% |
| 26 | | 00:08:16 | B | 11 | 0 | 5 | 0 | 0 | 0% |
| 27 | | 00:09:59 | B | 5 | 0 | 5 | 0 | 0 | 0% |
| 28 | | 00:09:25 | B | 5 | 0 | 6 | 0 | 0 | 0% |
| 30 | | 00:11:38 | B | 7 | 0 | 5 | 0 | 0 | 0% |
| 15 | No interaction | 00:05:50 | B | 0 | 0 | 0 | 0 | 0 | 0% |
| 32 | | 00:04:36 | B | 5 | 0 | 0 | 0 | 0 | 0% |
| 17 | | 00:01:41 | B- | 6 | 0 | 0 | 0 | 0 | 0% |
| 34 | | 00:05:49 | B- | 0 | 0 | 0 | 0 | 0 | 0% |
| 25 | | 00:04:05 | B- | 0 | 0 | 0 | 0 | 0 | 0% |
| 33 | | 00:03:59 | B- | 6 | 0 | 0 | 0 | 0 | 0% |

### *The interplay between interaction patterns and inquiry achievement*

Student interaction and interaction patterns were analyzed to reveal how different patterns of interaction affect inquiry achievement in the laboratory course. This study thus analyzes group interaction patterns: group performance as it relates to the five interaction types, time used to regulate group plans, and the ratio of conflict resolution to demonstrate how the students used PDAs in the laboratory classes to affect their achievement in the course. Because the PDAs were used to help students resolve conflicts and create shared inquiry plans, this study measures the resolution ratio within each group by formula (1). The formula can also be used to determine how different peer interaction patterns affect group inquiry achievement.

$$\text{Resolution ratio} = \frac{\text{Total number of resolution of options by group members}}{\text{Total number of changing of options}} \quad (1)$$

Table 2 shows the analysis of inquiry achievement and planning time in terms of peer interaction patterns and interaction activities. Students displayed different characteristics with regard to interaction patterns and interaction activities. Obviously, students required more planning time in concurrent informed cooperation and sequential informed cooperation than in concurrent concealed cooperation, because the former two patterns require time for students tounderstand each other' s opinions and to resolve conflicts. The average planning time for concurrent informed cooperative exploration and sequential informed cooperative exploration were 41 and 17 minutes respectively. In contrast, the average planning time of concurrent concealed cooperation was only 10 minutes.

Furthermore, groups whose interactions were classifiable as thesequential informed cooperation pattern showeda higher frequency of perceived conflicts than those demonstrating concurrent informed cooperation. Groups exhibiting concurrent informed cooperation also had a higher conflict resolution ratio than that of groups following the sequential informed cooperation pattern, probably because each student following the concurrent informed cooperation pattern, after independently revising the group plan, had acquired richer experience organizing group plans and thus was better prepared to establish agreement in later activity. Conversely, student groups adhering to the pattern of sequential informed cooperation, despite ultimately demonstrating high perceived agreement, were unable to achieve conflict resolution ratios as high as those of concurrent informed cooperation groups, possibly because students following other students' work when revising group plans lack personal experience in group plan development.

In the teacher' s evaluation of the final inquiry reports, the concurrent informed cooperation groups obtained the highest overall scores (ranging from A- to A+) of all interaction patterns. The groups using sequential informed cooperation achieved the second highest scores (ranging from B to A+). Table 2 shows that planning time and resolution ratio may each play an important role in affecting group inquiry performance. When using the patterns of concurrent informed cooperation and sequential informed cooperation, students spent more time planning and achieved higher conflict resolution ratios.

These patterns were defined on the basis of the various types of interaction engaged in by the groups. By comparing the differences and similarities of the 34 groups, some basic rules to categorize their interaction patterns could by developed:

- Concurrent informed cooperative exploration: Students spend a great deal of time planning their inquiry (41 minutes on average). Groups following this pattern demonstrated a high conflict resolution ratio (on average 61%) and high perceived agreement (7.4 on average). Student groups following this pattern achieved the highest overall scores (ranging from A- to A+) in the teacher's evaluation of the inquiry reports.
- Sequential informed cooperative exploration: Groups following this pattern showed a high frequency of perceived conflicts (10.3 on average). Students could perceive conflicts amongthemselves but did not reach a high conflict resolution ratio. Student groups following this pattern achieved the second highest overall scores (ranging from B to A+) in the teacher's evaluation of the inquiry reports.
- Concurrent concealed exploration: Students work independently with only limited interaction. Some conflicts go unperceived. Neither conflict nor agreement is perceived by students. Student groups following this pattern achieved the second lowest overall scores (B) in the teacher's evaluation of the inquiry reports.
- No interaction: There is no perceived conflict nor resolution. Student groups following this pattern achieved the lowest overall scores (ranging from B- to B) in the teacher's evaluation of the inquiry reports.

## 5. Conclusions

Because of the difficulties students encounter with inquiry in laboratory courses, this study attempts to provide students with regulative tools on PDAs to aid in regulating and interacting with peers. Analysis of student interaction through PDAs found five main types of peer interaction, i.e.comparison, perceived conflict, perceived agreement, conflict resolution, and unperceived conflict, when student cooperation is mediated by PDAs. The benefits of using PDAs in cooperative planning activities include the ability of PDAs to provide each student with free access to the artifacts (individual inquiry plans) produced by every group member and thus to negotiate directly with peers regarding those artifacts in order to produce group work. The four main interaction types (comparison, perceived conflict, perceived agreement, and conflict resolution) that students demonstrated through PDAs are important negotiation activities. This result reflects the effect PDAs have in facilitating the negotiation of group work.

However, unperceived conflicts were frequently observed because of the client-server nature of PDA software. The implementation of the regulation tools on the PDAs in this study did not enforce simultaneous synchronization among the PDAs of group members. In other words, the changes to the group plan made by one member were not immediately displayed on the PDAs of other members. Students were able to obtain the most recent state of the group work only by reloading after all members had submitted their latest group plans. This facet of the design resulted in the frequent occurrence of unperceived conflicts.Unperceived conflicts may lessen the ability of students to share their learning experiences with each other.Technological devices, PDAs in the case of this study, can record and reflect students' current ideas and states, but unperceived conflicts may still be concealed in the individual devices. Further investigation into the possibility of an effective status awareness mechanism to synchronize the work of individuals in PDAs may be required to alleviate the problem of unperceived conflict. For example, the reinforcement mechanism applied in MCSCL to maintain shared understanding among group members may be applied to establish mutual understanding.

This study also identifies four interaction patterns that are indicative of the different styles of interaction afforded by PDAs. The four interaction patterns are: concurrent informed cooperation, sequential informed cooperation, concurrent concealed cooperation, and no interaction. Teachers and system designers should take note of the advantages and disadvantages of using PDAs in facilitating group activities and use this information to avoid unfavorable interaction patterns such as concurrent concealed cooperation. Different interaction patterns may have different advantages and disadvantages. For instance, although students in groups using sequential informed cooperation could perceive the current status of members, they still could not easily resolve their conflicts. Such difficulties may result from the lack of prior experience in learning. On the other hand, student groups using concurrent informed cooperation could easily resolve conflicts but had problems with time constraints and some unperceived conflicts. However, the best way to help students efficiently negotiate and avoid unperceived conflicts still requires further investigation.

Students engaging in different interaction patterns showed different levels of inquiry achievement. Generally, the amount of time used to regulate group plans and the ratio of conflict resolution seems to affect student achievement significantly. A high conflict resolution ratio helps students achieve high scores in the final inquiry report. The study results confirm the importance of shared agreement in negotiation activities. This result is consistent with the current principle of collaborative learning, which states that students have to express and discuss divergent ideas in order to construct and share knowledge collaboratively (Puntambekar, 2006). Therefore, educators using PDAs in classrooms and the designers of groupware on PDAs should not only support students in expressing personal ideas but should also build a mechanism or a protocol, such as the display of different personal ideas as in this study or the MCSCL mechanism (Zurita & Nussbaum, 2004), to mediate the negotiation process and help establish shared agreement.

## References

Blumenfeld, P. C., Marx, R. W., Soloway, E., & Krajcik, J. (1996). Learning with peers: From small group cooperation to collaborative communities. *Educational Researcher*, 25(8), 37–42.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *The Journal of the Learning Sciences*, 8(3&4), 391-450.

Fry, J. M., & Hin, M. K. T. (2006). Peer coaching with interactive wireless technology between student teachers: Satisfaction with role and communication. *Interactive Learning Environments, 14*(3), 12.

Jonassen, D. H. (1994). Toward a constructivist design model. *Educational Technology*, 34(4), 34–37.

Hinrichsen, J., & Jarrett, D. (1999). Science inquiry for the classroom: A literature review. *The Northwest Regional Educational Laboratory Program Report*.

Hollingworth, R. W., & McLoughlin, C. (2001). Developing science students. *Australian Journal of Educational Technology, 17*(1), 50-63.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.

Littleton, K. & Häkkinen, P. (1999). Learning together: Understanding the processes of computer-based collaborative learning. In P. Dillenbourg (Ed.) Collaborative Learning: Cognitive and Computational Approaches (pp. 20-33). Pergamon: Oxford.

Liu, C. C., Chou, C. C., Liu, B. J., & Yang, J. W. (2006). Improving mathematics teaching and learning experiences for hard of hearing students with wireless technology-enhanced classrooms. American Annals of the Deaf, 151(3), 345-355.

Liu, C. C., Kao L. C. (2007). "Do Handheld Devices Facilitate Face-to-Face Collaboration?" : Handheld Devices with Large Shared Display Groupware to Facilitate Group Interactions, Journal of Computer Assisted Learning, 23(4), 285-299.

Liu, C.C., Chung, C.W., Chen, N.S., Liu, B.J. (in press). Analysis of Peer Interaction in Learning Activities with Personal Handhelds and Shared Displays. *Educational Technology and Society*.

Manlove, S., Lazonder, A.W., & Jong, T. (2006). Regulative support for collaborative scientific inquiry learning. *Journal of Computer Assisted Learning*, 22(2), 87-98.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? the case for guided methods of instruction. *The American Psychologist*, 59(1), 14-19.

Puntambekar, S. (2006). Analyzing collaborative interactions: divergence, shared understanding and construction of knowledge. *Computers & Education*, 47(3), 332-351.

Roschelle, J.(2003). Unlocking the learning value of wireless mobile devices. Journal of Computer Assisted Learning, 19(3), 260-272.

Stahl G., Koschmann T. and Suthers D. (2006) Computer-supported collaborative learning, in *Cambridge Handbook of the Learning Sciences*, ed. R. K. Sawyer (Cambridge University Press, Cambridge, UK).

Zurita G. & Nussbaum M. (2004) Computer supported collaborative learning using wirelessly interconnected handheld computers. *Computers and Education, 42*(3), 289–314.

# Examining Preservice Teachers' Ability to Attend and Respond to Student Thinking

Vicky Pilitsis and Ravit Golan Duncan, Rutgers University, New Brunswick, NJ 08901
Email: pilitsisv2@yahoo.com, rgduncan@rci.rutgers.edu

**Abstract:** In order to effectively implement inquiry practices in science, teachers must be able to attend to student thinking and adjust their lessons to build on and respond to student ideas. Research on experienced teachers suggests that their understandings of learners influence their instructional decision-making. However, the research on teacher education has mixed results about preservice teachers' ability to attend to students' ideas (Davis et al, 2006). Although preservice teachers may recognize that learners have prior knowledge they usually do not take into account students' ideas in their teaching practices (Friedrichsen et al, 2009). In this paper, we report on preservice teachers' ability to notice students' ideas (as manifested in written models) and respond to these ideas in subsequent instructional planning. Our data is drawn from clinical interviews conducted with 15 preservice teachers at the end of each of four consecutive methods courses (total of 60 interviews) in a two-year certification program.

## Introduction

Standards documents and recent policy reports have called for a nationwide emphasis not only on the content and processes of science but also on the epistemology and the practices of scientific inquiry (AAAS, 1993; NRC, 1996; Duschl, Schweingruber, & Shouse, 2007). Inquiry based instruction entails a student-centered classroom in which teachers attend to and work with students' ideas to develop models and explanations of natural phenomena. A key factor in implementing inquiry-based practices is the ability to attend to and interpret student ideas and to use that interpretation to guide instructional design (Hammer, 2000; van Zee & Minstrell, 1997; NRC, 1996; AAAS, 1993). Researchers have found that when teachers attend to student understanding and adjust instructional practices this can help minimize the gap between desired student performance and observed student performance (Bell & Cowie, 2001; Black & William, 1998). Davis, Petish, & Smithey (2006) conclude that although experienced teachers' understanding of learners influences their instructional decision making, the research on preservice teachers does not have any conclusive findings.

Some studies report that preservice teachers foresee few student learning difficulties when planning lessons (De Jong & van Driel, 2001; Simmons, 1999; Geddis, Onslow, Beynon, & Oesch, 1993). Other studies have found that preservice teachers do acknowledge the significance of students' ideas about science (Davis et al, 2006; Russell & Martin, 2007; Van Driel, De Jong, & Verloop, 2002). However, even when preservice teachers recognize that learners have prior knowledge they usually do not take into account students' ideas extensively in their teaching practices (Friedrichsen, Abell, Pareja, Brown, Lankford, & Volkmann, 2009). In a 1999 study, Tabachnik and Zeichner found that preservice teachers were interested in uncovering their students' understandings of particular concepts in science, but they did not take this information into consideration when designing their lessons. Moreover, Davis et al (2006) concluded that new teachers do not have very clear ideas about what to do with regard to students' ideas once they have surfaced them.

van Es and Sherin (2002) argue that reform based practices call for a new understanding of teaching and learning which requires teachers to develop new ways of interpreting the classroom. Furthermore, they claim that current teacher education programs do not focus on helping teachers to interpret interactions that take place in the classroom. In our research, we investigate the development of preservice teachers' ability to notice student ideas in written work, as well as the ways in which they subsequently respond to student ideas in their instructional plans.

## Theoretical Framework

According to van Es and Sherin (2002) the skill of noticing for teaching consisted of three aspects: (a) identifying what is important in a teaching situation, (b) making connections between specific events and broader principles for teaching and learning, and (c) using what one knows about the context to reason about the situation. The aspect of identifying what is important in a teaching situation refers to the ability of the teacher to select what they will attend and respond to in the class. Because there are so many different things going on in a classroom at any given point, the teacher must decide what is important and use this information to guide their instruction. Leinhardt, Putnam, Stein, and Baxter (1991) found that experienced teachers have "check points" and use these "check points" to judge how the lesson is going and decide how to proceed. Therefore, experienced teachers are more capable of recognizing and attending to what is important in the lesson (van Es & Sherin, 2002).

The next aspect of noticing, making connections between specific events and broader principles for teaching and learning, refers to the ability of the teacher to make connections between events that occur in the classroom and the broader ideas that they represent (van Es & Sherin, 2002). Similarly to the first aspect of noticing, as a teacher becomes more experienced, he or she will be able to determine how specific events relate to the process of teaching and learning. However, given the same situation, preservice teachers tend to only describe literal events that take place and miss making the connections to the bigger picture (van Es & Sherin, 2002). The last component of noticing, using what one knows about the context to reason about the situation, entails being able to take their knowledge of the subject and knowledge of how students think and apply it in the classroom. In our case its identifying relevant aspects of students' models, like accuracy of explanation or representation, explanatory power, use of evidence, and justification. The preservice teachers then evaluate these identified components to determine the students' understanding. Once the preservice teachers have evaluated the components in regards to student thinking than they determine what course of action is necessary, such as a class discussion or experiment, to clarify any misunderstandings.

In addition to learning how to notice, it is important for teachers to be able to take what is noticed and adjust their instructional practices. The activities that teachers undertake during instruction in order to produce information that can facilitate adaptation to on-going instruction are known as formative assessments (Sadler, 1989). Formative assessment can take many forms such as asking probing questions or class discussions, but it always includes a component of attending to and assessing student ideas in relation to the learning objectives and then adjusting instruction. In 1998, Black and Williams concluded that formative assessments improved student learning, especially for those students who were having difficulty with the content being taught. Formative assessment helps to prevent persistent student misunderstanding because the teacher is looking for evidence of student learning and adjusting instruction as needed (Erickson, 2007). Friedrichsen et. al (2009) found that teaching experience influenced the types and timing of assessments and that novice teachers did not readily implement formative assessments. For example, more experienced teachers tended to enact informal questioning throughout the lesson to check for understanding, while novice teachers waited to grade worksheets and only then revised subsequent lessons. Similar to the development of noticing, the ability to implement formative assessments are difficult and novice teachers struggle with making the connections between what students do not understand and what types of activities would promote student learning. In our study we are trying to help preservice teachers develop the competencies they would need to engage in the complex practice of formative assessment. The competencies they need to have are the ability to notice student thinking in written work and to design instruction that adequately responds to these interpretations.

A way to promote the development of noticing student thinking and designing instructional plans that correspond to these interpretations is through the analysis of student artifacts. The uses of artifacts of practice, such as video and student work have received a lot of attention in teacher education research (e.g. Kazemi & Franke, 2004). Several studies have advocated the use of student work as a tool for professional growth (e.g. Ball & Cohen, 1999; Little, 2002). Using student work has the potential to engage teachers in a cycle of experimentation and reflection and to shift teachers' focus from general pedagogy to one that is connected to their own students (Kazemi & Franke, 2004). However, empirical research on this topic is very limited (Kazemi & Franke, 2004).

In this paper, we hope to expand the research on the use of student work by examining the development of teachers' ability to notice and attend to student thinking as manifested in written artifacts. During clinical interviews, preservice teachers were given a task in which they were asked to analyze student models. We examined the development of teachers' ability to notice and interpret and respond (through instructional planning alone) to students' ideas in these student models. Specifically, our research questions are: (a) What do preservice teachers notice about student thinking as manifesting in student artifacts? (b) In what ways do preservice teachers attend to student thinking in their instructional design?

## Methods

### Study Context

This study was conducted in the context of a two-year Ed.M. certification program for secondary biology teachers at a large public university in the North East. The program included four life science specific methods courses that were taken in sequence. These courses were geared towards helping preservice teachers develop the knowledge and practices of inquiry-based teaching. The first course, *Methods I,* focused on the nature of scientific research and theory building. *Methods II,* the second course was essentially a design course in which the teachers worked in groups to design an extended inquiry based unit. *Methods III*, which accompanied the teaching internship, focused on the implementation of inquiry based lessons and on assessment strategies. Finally, the fourth course, *Methods IV,* engaged teachers in action research and the analysis of data collected during the teaching internship. All courses, especially the last two, included an emphasis on student thinking (including analysis of student work) and the importance of attending and responding to students' ideas. There were 15 preservice teachers who were enrolled in the program and completed all four courses in sequence.

## Data Collection and Analysis

At the end of each of the four courses we conducted clinical interviews with all the teachers. A faculty member and trained graduate students conducted the interviews. Although somewhat different protocols were used in each set of clinical interviews, there was a common student-model critique task, which is the data source analyzed for this paper. In this task the teachers were presented with three student models explaining how a cut heals (cellular division) or why ice floats (density); two versions were used for counterbalancing purposes. We created the student models to closely mimic student thinking and representations. The models consisted of a drawing with an explanation written underneath. Models varied in terms of level of details, scientific accuracy of explanation, use of evidence (from students' prior knowledge) and justification of how the evidence relates to the explanation provided by the model. The teachers were then asked to identify (a) good aspects of the models; (b) problematic aspects of the model; (c) description of their next instructional move if these were naïve models (done at the beginning of a unit); and (d) description of their next instructional move if these were revised final models (done at the end of a unit). Interviews lasted 30-45 minutes and were videotaped and later transcribed verbatim and blinded for analysis.

We conducted a content analysis and through an iterative process of constant comparison we developed a coding scheme to capture four emergent themes of the models the teachers critiqued. The four themes were: (a) model representation (e.g. labeling); (b) mechanism in model (e.g. model shows a before and after or progression); (c) use of evidence or justification in models (e.g. student supports belief with data); and (d) accuracy of ideas of the models. To capture the quality of the preservice teachers' critiques of the models, we focused on two aspects of the critique: (a) the specificity of information they provided and (b) the noticing aspect- what themes were noted by each teacher. In terms of specificity, the preservice teachers' responses were coded as one of three categories: (a) identification; (b) description; or (c) interpretation of student thinking. A statement was coded as identification if the teacher stated what was good or problematic about the model without providing any additional details about why that aspect of the model was good or problematic (e.g. "the explanation is good"). A comment was coded as a description if it involved both identifying an aspect of the model and describing what was positive or negative about this aspect (e.g. "the explanation is good because it contains some empirical evidence") . Finally, an interpretation comment involved identifying, describing, and interpreting a certain aspect of the model in regards to student understanding. A statement such as "the explanation is lacking because cells do not stretch in the healing process; the student is missing the concept of cellular division" would be coded as interpretation since the preservice teacher identified an aspect of the model (explanation), described the problem (cells do not stretch), and interpreted what this means in terms of the student's understanding (missing concept of cellular division). We also counted how many of the themes described above were discussed in the critique to obtain a value for the noticing aspect of our analysis. At the end of the analysis for the critique part of the model, each interview had a two part code (specificity code and a count of themes mentioned).

We then analyzed the last two parts of the interview task, in which the teachers were asked about their next instructional move given two conditions: students models were naïve or students models were final models. We developed a coding scheme that captured three aspects of these revisions: (a) attention to student ideas, (b) specificity of instructional moves, and (c) alignment with model based inquiry. The first aspect, attention to student ideas was coded as: (a) no mention of student ideas which was coded as a "1"; (b) mentioned specific students ideas in regard to a particular model which was coded as a "2"; and (c) mentioned specific student ideas and provided an instructional move that addresses these ideas which was coded as a "3" (e.g. "model #1 did not understand molecules so I would have the students draw out the molecular structure of water to show the polar bonds of water"). The second aspect, specificity of instructional moves, was coded as: (a) stated general activity with no further example which was coded as a "1"; (b) stated a specific activity which was coded as a "2"; and (c) stated a specific activity that has a logical progression in attaining the lesson objectives which was coded as a "3" (e.g. I would give the students data and have them work in groups to make a group model from this data and then present these models to the class and come up with a class model so that all the students are on the same page). The third aspect, alignment with model based inquiry, was coded as: (a) no alignment which was coded as a "1" (e.g. lecture, teacher centered questioning); (b) slight alignment which was coded as a "2" (e.g. demonstration, cookbook lab); (c) alignment which was coded as a "3" (e.g. collaboration with peers, argumentation).

## Results and Discussion

### What Preservice Teachers Notice in Students' Models

Our analysis of the teachers' critique of the student models, allowed us to determine what teachers noticed in students' work and in what ways they could interpret the models in relation to students' understanding. In terms of the aspects of students' work teachers attended to, we found that their comments pertained mostly to the mechanism and explanatory nature of the model (e.g. "model has no explanation"; "model shows wrong grain size"; "model has a before and after"). Comments that pertained to model representation were statements about the components that made up the model (e.g. labeling) and were the second most frequent theme. Finally, statements about the use of evidence and justification (e.g. "model is supported by data"; "student justifies their model") and accuracy of ideas in terms of the thoughts expressed in the models (e.g. "student's thinking about cells stretching is incorrect"; "student understands the concept of density") were less frequently noted (see Table 1).

We then examined how many different themes teachers noted in analysis of students' models. On average, we found that there were approximately 2.3 themes noted per interview; thus teachers, on average, discussed two such aspects of students' models. We found no difference in terms of whether their comments were on positive (e.g. "it is good that all the students know that ice does float") versus the negative aspect of students' work (e.g. "model #1 doesn't explain what happens to a molecule which is what the question is asking"). Although there was only slightly more themes mentioned in *Methods III*, overall the number of themes mentioned by teachers at the end of each course was the same. The preservice teachers tended to focus more on representational aspects of the models (e.g. labeling) in the first methods course and did not attend to the accuracy of the ideas or the use of evidence and justification. However, after *Methods II*, the teachers began to focus more on accuracy of ideas and use of evidence and justification.

Table 1: Frequency count of themes noticed in the students models by methods course

| Themes | Methods I | Methods II | Methods III | Methods IV |
|---|---|---|---|---|
| Mechanism in Model | 15 | 12 | 14 | 14 |
| Model Representation | 12 | 4 | 7 | 7 |
| Accuracy of Ideas | 0 | 9 | 8 | 7 |
| Use of Evidence and Justification | 4 | 6 | 8 | 7 |

Aside from analyzing what teachers noticed we also wanted to understand how teachers discussed what they noticed. Specifically, did they merely identify a problem (or leverage) in student thinking, or were they able to reason about, or explain, how what they noticed relates to student learning (for example, explaining what exactly the student confused about based on their model). We found that most of the teachers' comments were of the identification and description type (91 comments out of a total of 120) and that there were far fewer comments in which they elaborated upon and explained student thinking (29 comments out of a total of 120). Thus, the preservice teachers seemed to be more facile with describing what is problematic about a specific model, but were less able to explain what this means in terms of the student's state of understanding. However, it is the latter capability that is crucial to designing lessons that address students' ideas.

We also looked at whether teachers tended to provide more elaborate comments about positive or negative aspects of student models. We wanted to explore this to see if there was a significant difference between the good versus the problematic components noticed. There was a small, but consistent, difference between the number of identification comments made about positive versus negative aspects of the models (21 positive vs. 25 negative) and the interpretive comments made about the positive and negative aspects of the models (13 positive vs. 16 negative). However, we noticed that there were many more descriptions of positive aspects of the models provided by the teachers than descriptions about negative aspects (32 positive vs. 13 negative). Therefore, it seems that the preservice teachers were able to describe the positive aspects of the models more readily than the negative aspects, even though they generally did not notice many more positive aspects.

Our findings reveal that preservice teachers tended to focus on identifying and describing components of the student models rather than interpreting the models in regard to student understanding. In reference to van Es and Sherin 's (2002) components on the aspects of noticing, it seems that the preservice teachers are not making the connections between the specific events, or ideas, in the models and how that relates to the broader principles of teaching in learning. Our teachers did not seem to be attending to written responses as evidence of specific cognitive obstacles (or leverages) in student learning, rather many of them tended to analyze the models in a very literal and contextual manner. Additionally, it was evident that the teachers struggled to interpret what they saw in the models in relation to student understandings of the phenomenon. For example, one of the teachers realized that the model did not portray the phenomenon at the correct grain size, but was unable to see this was potentially due to a lack of knowledge about molecular structures. A plausible explanation for this trend was that the teachers may not have had strong subject matter knowledge about the modeled phenomenon (e.g. density and cellular division).

## Attending to Student Thinking in the Instructional Design

We next focused on whether and how the preservice teachers took into account and addressed student thinking in their instructional plans. In order to answer this question, we focused on the last part of the interview task in which the teachers were asked what they would do the next day in the classroom if the student models presented were (a) typical naïve models (created prior to formal instruction) and (b) typical final models (created after formal instruction). We analyzed three aspects of teachers' instructional designs: (a) attention to student ideas; (b) specificity of activities suggested; and (c) alignment with the model based inquiry pedagogical approach.

In regards to attention to student ideas, we gave each teacher a score, from 1-3, to reflect the extent to which they took into account student ideas in their plans. We found that, for the most part, teachers did not attend much to student thinking and that they did not substantially improve, in this regard, over time. The average scores for each course (for both naïve and final model questions) had only a slight difference (1.25-1.8). For example, during *Methods III* Sean was presented with the student models about why ice floats. He commented that the students were not using the correct grain size to represent the phenomenon that ice is less dense than water. When asked what he would do in class the next day if these were naïve models, he stated "I think the first thing I would do is ask them to collaborate with each other. Some have better ideas than others and after that perhaps the groups can do experiments themselves". While Sean felt that a major problem with students' models was that they were not at the molecular structure of ice and water (and hence their models were not at the correct grain size); his instructional move does not relate to this analysis in any specific way. The instructional move he suggested is a viable one (and we suspect that such a move may result in some discussion of molecular structures in some groups); however, it is a rather generic move and it doesn't take into account what he noticed about students' thinking.

In contrast, during *Methods II,* Catherine evaluated student models that focused on the healing of a cut. She commented that all the models had explanations but that some were incorrect. Catherine described her next instructional move as stemming from her interpretation of students' understandings: "Model #1 may understand that cells are involved [based on the written part of the model] but didn't incorporate it in the model. It is important that they know cells are involved so that is where I would start off". Catherine identified "cells" as the missing piece of the puzzle, and thus noted that her next instructional move would address the role of cells in the phenomenon. Unlike Sean, Catherine used her interpretation of students' understanding based on the models to inform her next instructional move, this sort of response was infrequent in our data set. Table 2 shows the teachers that received at least one score of "3" in regards to taking into account student thinking (a score of 3 corresponds to comments that mentioned specific student ideas and provided an instructional move that addresses these ideas). Out of the fifteen teachers, only eight of them received at least one score of 3. In comparing the different courses, the lowest average score (1.25) was during *Methods I* and the highest average (1.8) during *Methods IV.* We then calculated the overall average for all four courses for the instructional move after being presented with the naïve model and the instructional move after being presented with the final model, and found that the teachers took students' ideas into consideration slightly more when the plan was for the naïve versus final models (1.65 vs. 1.48). However, overall the teachers did not readily take into account student thinking when planning their next move in the classroom. This mirrors findings by Friedrichsen et al (2009) and Tabachnik & Zeichner (1999).

Table 2: Selected teachers that did attend to student thinking in their instructional design

| Teachers | Methods I | | Methods II | | Methods III | | Methods IV | |
|---|---|---|---|---|---|---|---|---|
| (pseudonyms) | Naïve | Revised | Naïve | Revised | Naïve | Revised | Naïve | Revised |
| Catherine | 2 | 2 | 3 | 1 | 3 | 1 | 2 | 2 |
| Anna | 1 | 2 | 3 | 1 | 3 | 2 | 1 | 2 |
| Ava | 1 | 1 | 1 | 3 | 1 | 2 | 1 | 3 |
| Patrick | 1 | 2 | 1 | 1 | 1 | 3 | 1 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Jackie | 1 | 1 | 3 | 1 | 3 | 1 | 1 | 1 |
| Nina | 1 | 1 | 1 | 3 | 1 | 1 | 2 | 1 |
| Jack | 1 | 2 | 3 | 1 | 3 | 3 | 3 | 2 |
| Claire | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 3 |

We next analyzed the specificity of their next instructional move; that is, how clear and specific were they about what they planned to do. We found that the last two methods courses had higher average specificity scores compared to the first two, with the highest average score of 2.07 for the instructional plan for final models in *Methods IV.* For example, after being presented with the student models about why ice floats during *Methods IV*, Bani discussed her next instructional plan for these naïve models as: "I would give them more data" a rather unspecified move; Bani didn't explain: (a) what type of data or (b) what she would have the students do with that data. However, Sean, after critiquing student models during *Methods III*, was much more specific with his instructional design. He stated that "I would have the kids get in small groups and share their models in small groups. I would then have the group create a consensus model and share them in front of the room and we would argue them". Sean's description of his next instructional move included information about the specific activities that would be performed in the classroom and these activities followed a logical progression.

We expected that the teachers may tend to be more specific about what they planned when confronted with "final models" because had the teachers attended to student thinking the specifics of the instructional designs would probably have related to specific challenges students' demonstrated in their models. Therefore, the activities being described in their next instructional move would be specific to the misunderstandings in the student model. For example, if the teacher critiques the models for not being the correct grain size, an activity that would be coded as a "3" for specificity would outline how doing X would improve student's understanding of grain size. However, there were very small differences (1.85 vs. 1.80) in terms of the specificity with which teachers described their next instruction in the context of the naive versus final models. This could be because the teachers did not readily attend to student thinking or understanding, and therefore, the specificity of the activities did not change during moves for the naïve versus final models because the teachers were not evaluating the models in regards to student understanding. Thus, the activities that the teachers described were most likely common activities, for that content, that they had implemented successfully before.

The last aspects of teachers' next-instructional-moves we examined were the alignment of these plans to the model-based inquiry approach which is the focus pedagogy of the teacher education program. This method entails a student centered classroom in which teachers attend to and work with students' ideas to develop models and explanations of natural phenomena. When looking at each of the four methods courses in turn, we found that during the first course the teachers' plans were not aligned with inquiry-based practices but were mostly teacher centered in nature (e.g. lecture). During the second and last course, the instructional plans were much better aligned with model-based inquiry practices (2.85 and 2.57 respectively). For example, during *Methods I,* Ava described her next instructional move as "I would talk about how a scab forms so that they [students] understand that the scab isn't doing the healing". Ava's practices are teacher centered and showed no connection to model based practices. However, the teachers' lessons began to align more with model based inquiry during *Methods II* and *Methods IV.* Patrick, during *Methods II,* stated that, "The best thing to do is have the students explain their models to the rest of the class and justify them. The class could work out the rest of the problems and allow them [students] to argue and allow other students to argue why that model would be incorrect". Patrick's solution of collaboration and argumentation are consistent with the principles put forth in the heuristic for progressive disciplinary discourse (HPDD) framework that aims to foster learners' participation in material and discursive activities characterizing the work of scientists (Windschitl, Thompson, and Braaten, 2008).

Comparison of the overall averages for the naïve-models instructional move and final-models instructional move, revealed that the teachers' plans were slightly more aligned with model-based inquiry practices for the naïve-models condition (2.36 average) versus the final-models condition (2 average). Thus the preservice teachers tended to suggest more model based inquiry practices when they were designing lesson activities assuming their students were just beginning to study the topic. For example, Molly discussed her instruction after the naïve models during *Methods I* as:

> "I would have a class discussion and everyone shares their naïve models to get everything out and then after the discussion we would do some sort of introduction or look at some kind of experiment. We will draw the raw models, discuss it, and then change the models after the class discussion and then have them [students] do some sort of experiment."

Implementing practices that Molly has outlined above, such as peer collaboration and interpreting data, are part of model-based inquiry instruction. However, for her instructional move given the final model, Molly stated "I would need to go back. I feel that I should go back and do some kind of activity or discussion with them even if it is a five to ten minute lecture". Molly, who previously advocated for student centered activities, now changed the focus of her instruction to lecturing." Molly's responses are not unique and in fact many of the teachers reverted back to teacher-centered activities after they assumed that an initial constructivist approach did not work. Grossman, Wilson, and Shulman (1989) concluded that teachers may revert back to the way they were taught if they are uncomfortable or uncertain of their abilities. Therefore, many of the teachers, including Molly, were unsure of other constructivist approaches to implement; thereby, reverting back to their comfort zone, teacher centered approaches.

In summary, our analyses suggest that the preservice teachers tended not to take students thinking into account when designing their next instructional practice, which could be due to a lack of content knowledge and/or pedagogical content knowledge. Without such knowledge, the teachers may not recognize problematic ideas expressed in the models, and they would not know what types of activities could help address the misconceptions they did identify in the student models. Erickson (2007) contends that teachers fail to make formative use of assessment data because the teachers do not know how to interpret the information to pinpoint alternative pedagogical "moves." Additionally, Davis and Smithey (2009) argue that further support is needed to help preservice teachers consider the factors that contribute to students' ideas or the resilience of those ideas, as well as to connect student ideas to instructional experiences. This resonates with our findings in this study. We did notice a slight increase in attention to student ideas during *Methods IV* which may be attributed to the action research activities teachers engaged with during this course. As part of their research projects teachers analyzed student artifacts from their class and had to relate their analyses back to their research questions and learning theories.

The instructional practices discussed by the teachers were often well aligned with model based inquiry practices. One important aspect to note is that after *Methods III,* the alignment to inquiry practices slightly decreased in comparison to *Methods II* and *Methods IV.* It was during *Methods III* that the preservice teachers were completing their student teaching practicum. Through several class discussions it was apparent that many mentor teachers did not implement these inquiry practices in their class and some were overtly against their use. It could be that some of the preservice teachers were influenced by this and may have begun to mimic the practices of their cooperating teacher. Therefore, this drop in inquiry practices could be contributed to the experiences they were facing during their student teaching, and perhaps a shift in their philosophical stance regarding effective pedagogy. Overall, during *Methods IV* showed improvements in all three categories- attention to student thinking; specificity of activities; and alignment with inquiry based practices. Although the improvement was small, it does show promise that action research projects do promote changes in preservice teachers' attention to student thinking in instructional design.

## Conclusions and Implications

Research has shown that beginning teachers typically focus on themselves as teachers, often at the expense of paying close attention to student learning (Meskill et al, 2002). Our findings support this assertion and illustrate the ways in which preservice teachers may begin to develop this skill. We believe that teacher education programs can encourage the shift from teacher (self) to student learning by engaging teachers in the practice of looking for evidence of student understanding in written artifacts as well as classroom discourse. In our methods courses we provide opportunities for teachers to evaluate student work and student discourse. However, our findings suggest that these opportunities were either not substantive or not frequent enough to engender the kind of evidence-based approach to instructional design we had hoped teachers would develop.

It seems that the single most influential course on teachers' ability to attend and respond to student thinking was the last methods course with its action-research focus. During this course, teachers analyzed various student artifacts for student understanding that they had collected during their student teaching internship the previous semester. In follow up studies, we would like to engage the teachers in analysis of student work and help them develop more robust strategies for looking at evidence of student understanding and relating such interpretations to potential instructional strategies.

## References

American Association for the Advancement of Science (1993). *Benchmarks for Science Literacy*, New York: Oxford University Press.

Ball, D. & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice based theory of professional education. In L.Darling-Hammond & G. Sykes (Eds.), *Teaching as a learning profession* (pp. 3-31). San Francisco: Jossey-Bass.

Bell, B., & Cowie, B. (2001). Formative assessment and science education. Dordrecht: Kluwer.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.

Davis, E.A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. *Review of Educational Researcher*, 34 (3), 607-651.

Davis, E.A., & Smithy, J. (2009). Beginning teachers moving toward effective elementary science teaching. *Science Education*, 93(4), 745-770.

DeJong, O., & Van Driel, J. (2001). The development of prospective teachers' concerns about teaching chemistry topics at a macro-micro-symbolic interface. In H. Behrednt, H. Dahncke, R. Duit, W. Graber, M. Komorek, A. Kross, & P. Reiska (Eds.), *Research in Science Education: Past, Present,and Future.* Dordrecht: Kluwer.

Duschl, R.A., Schweingruber, H.A. & Shouse, A. W. (Eds.) (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8.* Washington, DC: National Academies Press.

Erickson, F. (2007). Some thoughts on proximal formative assessment of student learning. In P. Moss (Ed.), *NSSE Yearbook.* Hoboken: Wiley.

Friedrichsen, P., Abell, S., Pareja, E., Brown, P., Lankford, D., & Volkmann, M. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching*, 46, 357-383.

Geddis, A.N., Onslow, B., Beynon, C., & Oesch, J. (1993). Transforming content knowledge: Leaming to teach isotopes. *Science Education*, 77, 575-591.

Grossman, P., Wilson, S., & Shulman, L. (1989). Teachers of substance: Subject matter knowledge for teaching. In M.C. Reynolds (Ed.) *Knowledge base for the beginning teacher* (pp. 23-36). Oxford: Pergamon Press.

Hammer, D. 2000. Teacher inquiry. In J. Minstrell & E. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 184-215). Washington, DC: American Association for the Advancement of Science.

Kazemi, E., & Franke, M.L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education*, 7, 203-235.

Leinhardt, G., Putnam, R.T., Stein, M., & Baxter, T. (1991). Where subject knowledge matters. In P. Peterson, E. Fennema, & T. Carpenter (Eds.). *Advances in research on teaching.* Greenwich: JAI Press.

Little, J. (2002). Locating learning in teachers' community of practice: Opening up problems of analysis in records of everyday work. *Teaching and Teacher Education*, 18, 917-946.

Meskill, C., Mossop, J., DiAngelo, S., and Pasquale, R. (2002). Expert and novice teachers talking technology: Precepts, concepts, and misconcepts. *Language Learning and Technology*, 6, 46-57.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academic Press.

Russell, T., & Martin, A.K. (2007). Learning to teach science. In S.K. Abell & N.G. Lederman (Eds.), *Handbook of Research in Science Education.* Mahwah: Lawrence Erlbaum Associates.

Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.

Simmons, P.E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D. Richardson, L., Yager, R., Craven, J., Tillotson, J., Brunkhorst, H., Twiest, M., Hossain, K., Gallagher, J., Duggan-Haas, D., Parker, J., Cajas, F., Alshannag, O., McGlamery, S., Krockover, J., Adams, P., Spector, B., LaPorta, T., James, B., Rearden, K., & Labuda, K. (1999). Beginning teachers: Beliefs and classroom actions. *Journal of Research in Science Teaching*, 36, 930-954.

Tabachnick, B.R., & Zeichner, K.M. (1999). Idea and action: Action research and the development of conceptual change teaching of science. *Science Education*, 83, 309-322.

Van Driel, J., DeJong, O., & Verloop, N. (2002). The development of preservice chemistry teachers' pedagogical content knowledge. *Science Education*, 86, 572-590.

van Es, B., & Sherin, M. (2002). Learning to notice: scaffolding new teachers' interpretations of classroom interaction. *Journal of Technology and Teacher Education*, 10, 571-596.

van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *Journal of the Learning Sciences*, 6 (2), 227-269.

Windschitl, M., Thompson, J., & Braaten, M. (2008). How novice science teachers appropriate epistemic discourses around model-based inquiry for use in classrooms. *Cognition and Instruction*, 26, 310-378.

# Changes in Teachers' Ability to Design Inquiry-Based Lessons During a Two-Year Preparation Program

Augusto Z. Macalalag Jr., Ravit Golan Duncan, Rutgers University, New Brunswick, N.J.
Email: azmjr@eden.rutgers.edu, ravit.duncan@gse.rutgers.edu

**Abstract:** Current reforms require teachers to design effective inquiry-based lessons. This is a challenging task particularly for preservice teachers who may not have experienced inquiry learning, and who do not posses a large repertoire of teaching strategies, or knowledge of student thinking in the domain. Here we report on the development of preservice teachers' lesson designs in the context of a two-year certification program with four consecutive methods courses. These courses included multiple opportunities to plan and implement inquiry-based lessons and units. We analyzed the lesson-designs of 15 preservice teachers. These lessons were generated as part of clinical interview conducted at the end of each course. Analysis of lessons revealed growth in the teachers' ability to: (a) craft appropriate questions to gather students' preconceptions and drive inquiry lesson, (b) anticipate students' prior knowledge and attend to the broader context of the curriculum, and (c) create more relevant investigations.

## Introduction

Current reforms advocate teaching scientific inquiry, which "refers to the diverse ways in which scientists study the natural world and propose explanations based on evidence derived from their work" (National Research Council, 2000, p.1). Teaching science through inquiry is congruent with constructivist perspectives of a more student-centered approach that promotes the learning of both science content and practices in a social setting (Anderson, 2007). However, over the past several decades, science educators and researchers have used and defined inquiry differently in their classrooms, curricula, and research projects. Historical definitions of inquiry in the classroom have ranged from traditional hands-on laboratories to open ended student-driven investigations (Windschitl, 2004). In this study we define model-based inquiry teaching as a form of inquiry that emphasizes the role of models in scientific practice, in particular, the use of models to build, revise and argue about scientific knowledge. Successful teaching of inquiry, model-based or otherwise, is challenging for most teachers particularly preservice teachers (Schwarz & Gwekwerere, 2007; Windschitl & Thompson, 2006). In a study of preservice elementary teachers, Hayes (2002) found that preservice teachers struggled in their new roles as teachers of inquiry. Specifically, he uncovered three major difficulties: letting go, going with students' interests, and asking the right questions. It was challenging for the teachers to let go of the didactic approach to teaching and move toward more student-centered instruction.

Teaching is a complex process that involves conceptualization of the teacher's intent and then the execution of the plan in the context of the classroom given the particulars of student ideas and responses to the plan. Lesson plans reflect teachers' thinking and the multiple decisions that teachers make before actual instruction begins (Duschl & Wright, 1989). While lessons plans do not usually reflect the nuances and complexities of implementation, they can provide a reasonable picture of the ways in which knowledge of inquiry is applied to instructional design. In this paper, we examine the ways in which preservice teachers' ability to plan (but not implement) inquiry-based lesson designs changes over the course of four consecutive science Methods courses in a two-year preparation program.

## Theoretical Framework

Lesson planning and design are windows to teaching practices. Planning refers to teachers' conceptualization and formulation of courses of action in a lesson, which has a profound influence in teachers' classroom behavior and students' learning (Shavelson, 1987). In planning and preparation, teachers demonstrate their knowledge of content and pedagogy, knowledge of students, selection of instructional goals, knowledge of resources, design of coherent instruction, and assessment of student learning (Danielson, 1996).

Several studies have explored aspects of experienced teachers' lesson planning process. Peterson, Marx, & Clark (1978) investigated the relationship between teacher planning, teacher behavior, and student achievement. The study focused on 12 experienced teachers who taught social studies lessons to three groups of high school students. Findings from the analysis of planning statements of the teachers indicated that the largest portion of planning focused on the content and selecting activities to be taught. A comparably small number of planning statements concerned the materials and the learner, and a very small number of planning statements mentioned lesson objectives. Brown (1988) conducted a case study of 12 middle-school teachers' yearly, unit, weekly, and daily planning. Analysis of written plans, think-aloud, and a questionnaire, indicated that the common factors that affect teachers' planning included student ability (very often), district curriculum guides, orderly transition between activities, student attention, standardized tests, and undergraduate training (rarely). Brown (1993) conducted a longitudinal study of two novice secondary teachers' instructional planning. In this study, she saw that cooperating teachers' planning practices, university professors' classes, textbooks, and curriculum guides influenced lesson planning. Duschl and Wright (1989) investigated the teachers' decision making models for planning and teaching of science of 13 high school teachers. They found that selection, planning, and designing of instructional tasks were dominated by considerations about student development, objectives set by the curriculum, and accountability pressures. However, teachers paid little attention to the scientific theories involved. These studies pertain to in-service teachers and highlight the complexity of the lesson design task and the different factors that influence teacher decision-making and planning.

The state of affairs regarding preservice teachers' lesson planning is similarly complex. Several research efforts to develop preservice teachers' ability to design lessons, specifically inquiry-based, have met with mixed success (Friedrichsen, et. al., 2009; Schwarz & Gwekwerere, 2007). Friedrichsen, et.al, (2009) analyzed the lessons developed by four teachers in a certification program: two interns and two full time (alternative route) teachers, to investigate differences between the two groups. Their analysis of the lessons revealed that both groups relied primarily on their subject matter knowledge and general pedagogical knowledge to plan the lessons. Both groups of teachers lacked topic specific knowledge about learners, instruction, curriculum, and assessment. Friedrichsen and colleagues found that a typical lesson for both groups began with the teacher asking questions, followed by a lecture and guided practice designed to memorize and practice the lecture material; these lessons were mostly teacher centered. Their research suggests that preservice teachers, in either route, do not tend to take students' prior knowledge and curriculum into account when designing lessons, nor do they have the necessary pedagogical content knowledge to address specific ideas that students may have about the content.

On the other hand, research by Schwarz and Gwekwerere (2007) suggests that by using highly scaffolded frameworks for instructional design, preservice teachers can begin to develop lessons that are more reform oriented. Schwarz and Gwekwerere (2007) used a guided inquiry and modeling instructional framework (termed EIMA) to support K-8 preservice science teachers in developing lesson plans and units. The preservice teachers learned and used the EIMA framework for their instructional activities and lesson planning in one science method course. Their findings suggest the framework was successful in helping teachers increase the use of different models to engage students, and move toward a more reformed based approach to teaching (i.e. conceptual change, inquiry, and guided inquiry). However, at the end of the semester, these researchers found that the preservice teachers still struggled with the concepts of scientific models and modeling.

In a different study, Windschitl, Thompson, and Braaten (2008) used the Heuristics for Progressive Disciplinary Discourse (HPDD) framework to improve preservice teachers' epistemic discourses in science. Their research showed that while most preservice teachers improved their knowledge of the function and nature of models, they failed to incorporate it into their inquiry lessons. Windschitl et al., (2008) thus showed that with appropriate scaffolding preservice teachers could improve their concepts of scientific models, concepts that Duschl and Wright (1989) had previously found to be challenging for teachers. These studies highlight the challenges that preservice teachers face when trying to use inquiry approaches in their instructional design.

In our Methods courses, preservice teachers engaged in designing, revising, and implementing inquiry based lessons in four consecutive courses. The intensive focus on lesson design was part of Methods II and Methods III, which are described in the next section. We anticipated that our preservice teachers would increase their attention to students' learning, curricula, and scientific models in their lesson plan and design over the four courses. More specifically, we hypothesized that the initial lesson designs would focus on the selection of activities from curricula and textbooks in ways that are similar to what Peterson, Marx, & Clark (1978) found in their study. Given the findings from the studies by Brown (1988) and Duschl and Wright (1989), we predicted that lessons would begin to focus on students' learning and development. Focusing on student ideas is challenging for teachers to do but is central to the constructivist approach. In contrast to Duschl and Wright's (1989) findings, however, we hypothesized that our preservice teachers might be better at focusing on scientific knowledge building (models and theories); ideas that are at the core of scientific inquiry.

## Methods

### Study Context

A qualitative approach was used to understand the extent to which the preservice teachers' ability to design lessons change over time. The participants are fifteen (4 male and 11 female) preservice teachers enrolled in a two-year biological science teacher education/certification program in a large university in the north east of the U.S. All preservice teachers have at least 30 credit hours in Biological Sciences before entering the program. Specifically, one teacher has a PhD in Biology, four have Bachelors of Science in Biology, and ten are in their forth year of a five-year Biological Sciences or related (e.g. Animal Sciences, Ecology and Natural Resources, and Environmental Science) programs. Moreover, two teachers have extensive research experiences, one as a senior scientist and the other as a researcher/laboratory manager, in commercial laboratories and another two worked as research assistants at the college during their undergraduate programs. As part of the certification program, students completed four subject-specific Methods courses in consecutive semesters. All courses were taught by one professor, the second author of this paper. In the first method course, Methods I, preservice teachers were engaged in science inquiry activities, readings, and discourse that promoted their understanding of scientific inquiry and theory development. During this course, the teachers, as a class, developed a framework for model-based inquiry that was to inform their lesson development in the subsequent course. The second course, Methods II, was a design-based course in which the teachers, in small groups, developed extended model-based inquiry units about selected topics in biology. In this course, teachers were introduced to several design frameworks including Learning for Use (Edelson, 2001) and Backwards Design (Wiggins and McTighe, 1999). At the end of this course they developed and implemented a single model-based inquiry lesson in their observation placement classrooms. The third course, Methods III, was a weekly seminar associated with their 15-week student teaching internship. Teachers were placed in local middle and high schools with teachers who may or may not have been familiar with reform-oriented science teaching. In this course, the teachers planned and implemented numerous lessons. They were required to extensively reflect on two short inquiry based units that they developed and implanted in their student teaching placements. In the final course, Methods IV, teachers conducted action-research projects aimed to develop their skills as reflective practitioners. Teachers used data collected during their student teaching to answer practical questions related to the use of the model-based inquiry approach in the science classroom.

## Data Sources and Analysis

We conducted clinical interviews with each teacher at the end of each of the four Methods courses. The interview protocol had four tasks that included defining model-based inquiry, critiquing a lesson, designing a lesson, and evaluating students' written work examples. In this report we focused on the analysis of the third task: the lesson design. We created two comparable versions of this task for counterbalancing purposes and each version was alternated across the four end-of-course interviews. Version A had three objectives for the topic of photosynthesis, while version B had similar objectives for the topic of cellular respiration. The teachers were given the objectives, asked to plan a lesson or short set of lessons that would address the objectives given to them, and then prompted to explain their lesson. This design task lasted for about 5-10 minutes.

Data analysis for this task began with the identification of essential features of inquiry-based lessons (e.g. hook, procedures, investigations, and assessment) and the systematic description of interrelationships among these features (Wolcott, 1994). The development of our coding schemes proceeded through an iterative process of application to the data set and refinement of the codes to capture relevant emerging themes in the data (Corbin & Strauss, 2008; Merriam, 1998). The first coding pass gave us a list of the different activities in the lesson (e.g. questions, investigations, student modeling, etc.) that teachers described in their interviews. We then examined these to identify patterns of change.

Through recurrent comparison of transcripts from interviews at different points in time we were able to identify shifts in the nature and quality of teachers' lesson designs with regard to several dimensions: (a) specificity and suitability of teacher *questions*, (b) *student-centeredness* of lesson*,* and (c) development of *investigations.* We will describe these dimensions in detail in the next section. Within the first dimension, we counted the specificity and suitability of the question to drive the need to know and students' prior knowledge in the lesson. For example, the following questions were constructed by teachers to drive a lesson on photosynthesis: (1) "how do plants grow, what do they need, what are they made up of, what are they made from" and (2) "I pose a question to the students on how do plants and humans work together in order to survive." In the first question, the teacher proposed specific questions to gather students' preconceptions about photosynthesis using plants' growth, sources of energy, and its composition. On the other hand, the second question was too broad and less suited to drive the lesson on photosynthesis. Students will most likely mention plants as a source of food and shelter to answer the question, which are correct but beyond the topic of the lesson. Moreover, the question did not clearly reflect the teacher's intention of using the question to deliver concepts such as human consumption of oxygen from plants and plant use of carbon dioxide from humans.

Within the second dimension, we coded the number of lessons that were procedural or activity oriented. These lessons have a set of activities that were predetermined by teachers, have no indication whether students' preconceptions will change the lesson, and reflected teacher-centered approach of teaching. We also coded the number of lessons in which teachers' voiced students' ideas, for example, Nadia mentioned during her planning that "[students] understand [energy being stored] and in carbohydrates, not only do they get how the molecule is being synthesized but they also get why for the energy and this energy is stored in molecular bonds," which indicated her sensitivity to students' knowledge. Furthermore, within the first dimension, we coded the number of lessons in which there was attention to the broader context of curriculum and to students' prior knowledge. These lessons contained careful sequencing of activities with respect to science content and scaffolding of students' knowledge.

Lastly, within the *investigations* dimension we counted how many investigations were open *or unstructured* (e.g. investigations that asked students to create their own experiments with minimal directions from the teacher). We counted how many of the lessons had investigations that provided specific data to the students, this included investigations using experimental data from a published science experiment or purposeful teacher constructed data.

## Results

### Specificity and suitability of Teachers' Questions

Most teachers mentioned *questioning* as one of the essential parts of a science lesson. These questions were often posed to students in the beginning of the lesson to serve as a motivator or hook, to guide the activities in the lesson, and to ascertain students' preconceptions about the topic. In analyzing the questions that the teachers created for their lessons, we found that lesson questions from the interviews of the third and fourth Methods courses were more specific and suitable for the topic than the questions generated from the interviews in the first and second Methods courses. In other words, teachers' seemed better able to form more specific and more suitable questions as the courses progressed (see Figure 1).

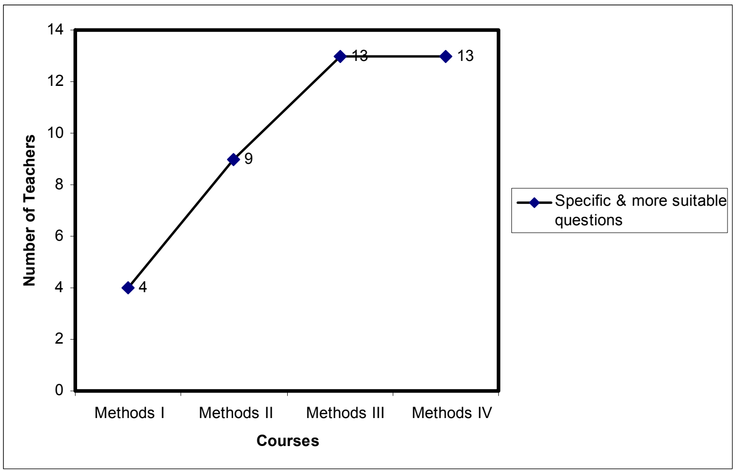



Figure 1. Specificity and suitability of Teacher Questions

Clare's questions from the second and fourth Methods courses are provided below as illustrative examples. In both cases Clare was designing a lesson on photosynthesis and her questions were posed at the beginning of the lesson.

> First, have a discussion with the class and maybe a hand-out about the purpose of photosynthesis and maybe what the students think what the purpose is and umm call out students to write on the board the main ideas and the key players in this process (Clare, Methods II)
>
> [Ask students] how plants get their energy and see what generates from that. (Clare, Methods IV)

Clare's question in Methods II is problematic in that it requires that students know the term photosynthesis and the idea of a biological process as having a purpose is somewhat teleological. It is unlikely that such a question would yield substantive participation from students; nor would it reveal relevant prior conceptions that, while incorrect, may be suitable building blocks (as students may not associate these ideas with the scientific name for the process). Her second question is couched in terms that students are likely familiar with (energy) and is more likely to enlist broader class participation and useful insights about students' prior knowledge. While the first question involves a more specific term (photosynthesis), it is the second question that specifically deals with the objectives of the lesson, namely, understanding the process by which plants get their energy.

### Student-Centeredness of Lessons

We also found shifts in the extent to which the designed lessons were student-centered. These changes included a decrease in the number of procedural or activity-focused lessons, an increase in voicing students' ideas in the lessons, and an increase in attending to broader context of the curriculum and prior knowledge of students (see Figure 2).

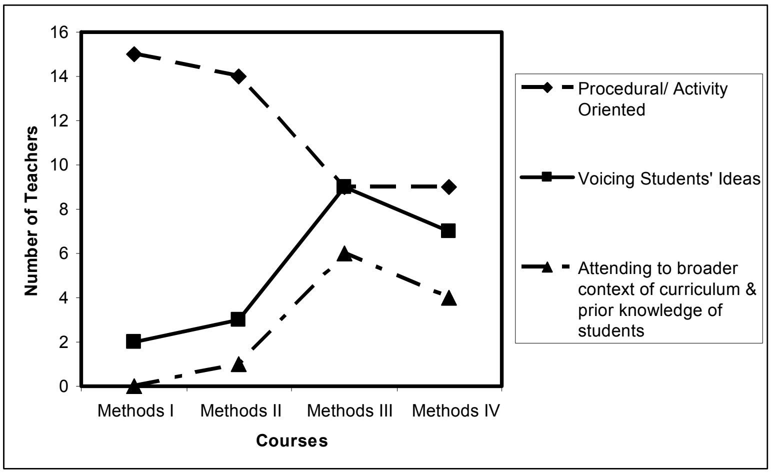



Figure 2. Student-Centeredness of Lessons

Specifically, almost all the lessons (93% to 100%) that the teachers created in the first two courses were procedural or activity-driven. For instance, Jackie's lesson about cellular respiration was primarily teacher-driven with students engaging in some inquiry practices, like model building, but, even then, they work with existing literature.

> Well, I think I would start with just a brief introduction of what cellular respiration is. And then I would have the students make a naïve model based on this pathway, like the overview that I give them of cellular respiration. And, in the model, I would want them to explain…[pause] from there I would probably give them literature. So then I would have them make – revise their naïve model based on that information. They will have to read about it, they will talk about it in small groups, possibly of two to three individuals in a group… discuss it, and make up a new model. (Jackie, Methods I)

Although Jackie included models and modeling in her lesson, there was no indication of how the students' naïve models or revised models will influence the course of instruction or the sequence of activities in the lesson. The majority of lessons in interviews from Methods I and II were similarly teacher-centered. In many cases, the lesson was an amalgam of activities that the teacher was familiar with (often from their own experiences as learners) without a clear connection to the lesson objectives and without any reference to students' expected level of understanding or the broader curriculum context. In addition, while the teachers described activities and practices that were inquiry-oriented, there was little consideration of students' prior knowledge or how the students' models/explanation could influence the lesson.

However, in the lessons designed during the interviews of Methods III and IV, 46% to 60% of teachers voiced-out students' ideas in the lesson and they began to attend to broader context of the curriculum in terms of what students might have already learned (and what prior knowledge they may have as a result). By "voicing-out" students' ideas, we mean that the teachers would speculate what sorts of responses they may get from students, often acting out student talk. Teachers also paid greater attention to potential difficulties that students may have in an activity. Christine's lesson exemplified this case. She mentioned possible conceptions or ideas that her students may have regarding where plants get their food.

> I would start by asking how plants get their food if they don't get it by eating things. The kids are going to be like "Venus flytrap eats flies" but that is the exception to the rule. So kind of get kids to think about how plants are stationary so how are they able to get food. See if anyone knows that plants make their own food, so ask them to explain a little more along those lines. But I am pretty sure that kids in high school don't have a grasp on photosynthesis so umm after that, say let's take a look at things that plants do have access to, like sunlight, air, nutrients, soil, water and then, from there, break down air because it is a mix of gases. Get them to kind of isolate carbon dioxide plants make oxygen. (Christine, Methods III)

Christine's lesson is more student-centered compared to Jackie's lecture-based lesson and she actually voiced out what students might say in response to her questions (Venus flytrap eats flies), and what knowledge they may have that she can build on. In another example, Catherine viewed her line of questioning as contingent on students' conceptions.

> First, I would have them do an initial model of, "how do animals get their energy?" And that would just be so they can get their ideas out there, and I can also see what they're thinking about this. I'm hoping that one of them would say, "food," because then, maybe I, with enough questions, I could pull out of them, "glucose." I'm assuming they've already learned about plants, with photosynthesis. And, so maybe I'd give them a data sheet, with information about …. [pause] or maybe my line of questioning might be, well, "what do chloroplasts store?" and try to get them along that line of thinking so maybe I can get them to come around and say, like, "oh, I guess maybe humans or other animals need glucose for energy." (Catherine, Methods III)

Like Catherine and Christine, teachers became not only cognizant of the students' prior knowledge and experiences but also willing to change the lesson based on what students brought to the lesson. Lastly, a number of teachers (26% to 40%) began to attend to the broader context of curriculum. This included consideration of the sequence of topics in a unit (or over a semester) and references to prerequisite knowledge that students must have prior to the lesson. In the following example, Nadia mentioned the concepts that her students learned before the lesson, the knowledge that they have to contribute in the lesson.

> Of course how I start it depends on what they learned before that. And I just took an example that I had in school. We learned carbohydrates before that and I actually did that on purpose. I did the bio-molecules before and I would do that again because it is a nice flow. Throughout the bio-molecules, they get the idea about the energy being stored… So they have this previous knowledge already and they know plants do that at least high school, they know plants do photosynthesis and build sugars and they mostly know that, at least in high school that is the case. So it is not hard to ask the connection. (Nadia, Methods III)

Nadia's teaching experiences in Methods III helped her reflect on the sequence of topics and what students are expected to know based on that sequence.

## Nature of Investigations

Another dimension of change was the teachers' ability to design *investigations* for their students. These changes included a decrease in open or unstructured activities, and an increase in providing data to students (see Figure 3).

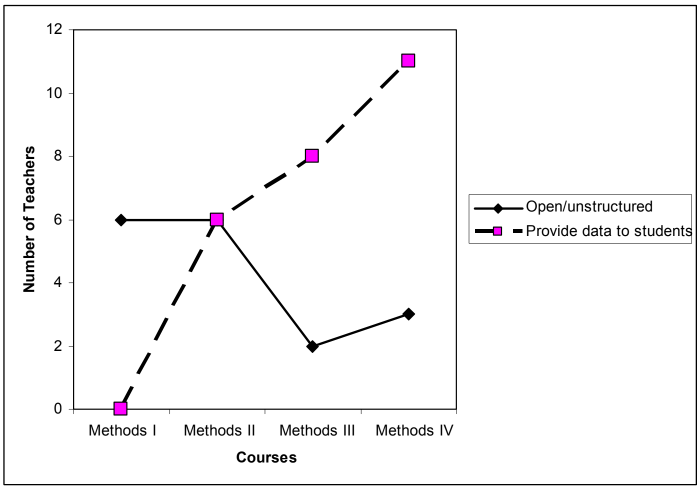



Figure 3. Development of Investigations

Less than half (40%) of teachers in Methods I and II proposed investigations that were open-ended or ill-structured. For example, Jake proposed a discovery-type of investigation in which the students, rather than him, would design and execute relevant experiments.

> I am trying to think of something more specific [investigation] but I can't think of it. I mean in general I would have them do that [experiments with carbon dioxide and oxygen] and it wouldn't necessarily be something that I designed. But I mean I would definitely give them the tools and let them play around with it and see to see what they were able to come up with. (Jake, Methods I)

Interestingly, we noticed a decline in open-ended investigations in the interviews from the last two courses. We believe that this may be due to teachers' realization (from to their experiences during their student teaching internship) of how challenging it is to implement open or unstructured investigations in the classroom. Instead of open-ended investigations, teachers tended to suggest investigations where teacher provides data to students or teacher-guided investigations.

In terms of providing data to students, we found that after Methods II, more than half of the teachers (60%) suggested investigations that involve providing data to students. Jake in Methods III, chose to provide data to his students, instead of giving them an open-ended problem, to address the same concept of energy in photosynthesis. Jason mentioned the data he would provide, the purpose of activities and the target concepts of the lesson. He also used an authentic problem context and data to motivate students to learn the material (the fertilizer runoff problem).

> Then I would give them some data that when there is fertilizer runoff, the amount of algae in the lake increase and the fish are still alive. But as the amount of algae kept going up and the fish are still there… but as the algae increased some more, fish eventually drops off. And I would have them revise their model based on this and I would anticipate that the first one will say something about the fertilizer and the second one will be the algae that somehow killed them. And I would have them either ask them to do an experiment that when you put a plant in the light and one in the dark and put a pH indicator and blow in it, using bromothyl blue in it, and the light will filter back to blue cause it took the $CO_2$ out and the dark it will turn the same or yellow to show that there is more $CO_2$. (Jake, Methods III)

Providing data to students or finding data for students to look at allowed teachers to focus on concepts that they wanted their students to learn, reduce open-ended or ill-structured activities, and promote other scientific practices such as analyzing data and modeling.

## Conclusion and Implications

The purpose of this study was to examine the changes of teachers' ability to design lessons as they progressed through a teacher certification program. Based on our analyses of the lesson design task in their end-of-course interviews, we characterized several dimensions along which there seemed to be shifts in teachers' abilities. The first was an increase in their ability to craft specific and better suited questions to gather students' preconceptions and drive the inquiry activities in their lessons. In the initial interviews, teachers' questions were typically too broad, which mirrors findings of Friedrichsen, et.al. (2009). However, the teachers in our study, given its longer duration, developed more specific and more suited driving questions in the last two Methods courses.

We also identified a trend of increased attention to student thinking and the boarder curriculum, and the use of more student-centered instructional approaches. In the first and second Methods courses, our teachers developed lessons based on selected activities similar to what Paterson, Marx, and Clark (1978) saw in study of their twelve experienced teachers. However, in Methods III and IV, our preservice teachers started to attend to students' prior knowledge and larger curricular context in designing their lessons. They started to voice-out students' ideas in the lesson and they began to attend to broader context of the curriculum in terms of what students might have already learned. This was similar to what Duschl and Wright (1989) found in high school science teachers' attention to student development in their lesson planning and instructional decision making, and what Brown (1988) found in the lesson planning of middle school teachers. Moreover, the progression to a more student-centered lesson and increased attention to the broader context of the curriculum suggests an increase in teachers' knowledge of learners, knowledge of context, and pedagogical content knowledge (Abell, 2007; Friedrichsen, et.al, 2009). The last dimension of changes was found in the teacher-developed investigations. These changes included a decrease in open or unstructured activities and an increase in providing data to students. We believe that the decrease of unstructured activities was due to the teachers' realization of how challenging it is to implement his kind activity. The shift of providing data to students allowed teachers to focus on concepts that they wanted their students to learn and on other scientific practices (e.g. modeling and data analysis).

In terms of scientific inquiry in the lessons, we saw that preservice teachers used the strategy of having students create naïve models (prior to any formal instruction) to gather their pre-conceptions about the question and designed investigations based on student learning. The model-based inquiry practices were congruent with the efforts of Schwarz and Wright (1989) and Windschitl, Thompson, and Braaten (2008). Analysis of Methods courses showed that after Methods II, which focused on instructional design, we noted increases in specificity and suitability of teachers' questions, and an increase in the number of lessons in which teachers provided students with data to analyze. After Methods III and IV we observed increases in student-centeredness of lessons (voicing of students' ideas and attending to broader context of the curriculum). Lastly, after Methods II, we saw a decrease in unstructured activities and an increase in providing data to students as part of investigations. We believe that changes in Methods II were due to the extensive focus on instructional design and multiple design frameworks in this course and the changes in Methods III occurred due to teachers' teaching experiences, interaction with students, and analysis of students' works. This latter development is congruent with the shift in focus from self (perceptions of one's capabilities as a teacher) to a focus on students and learning (Darling-Hammond & Baratz-Snowden, 2005).

## References

Abell, Sandra K.. (2007). Research on Science Teacher Knowledge. In Abel, S.K. & Lederman, N.G. *Handbook of Research on Science Education.* Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers.

Anderson, Ronald .(2002). Inquiry as an organizing theme for science curricula. In Abell, S.K. & Lederman, N.G. *Handbook of Research in Science Education.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Brown, D.S. (1988). Twelve Middle-School Teachers' Planning. *The Elementary School Journal, 89*(1), 69-87.

Brown, D.S. (1993). Descriptions of Two Novice Secondary Teachers' Planning. *Curriculum Inquiry, 23*(1), 63-84.

Corbin, J. & Strauss, A. (2008). *Basics of Qualitative Research, 3rd ed.* Thousand Oaks, CA: Sage.

Danielson, Charlotte. (1996). *Enhancing Professional Practice: A Framework for Teaching.* Alexandria, V.A.: Association for Supervision and Curriculum Development.

Darling-Hammond, L. & Baratz-Snowden, J. (2005). A good teacher in every classroom. San Francisco, CA: Jossey Bass.

Duschl, R.A. & Wright, E. (1989). A Case Study of High School Teachers' Decision Making Models for Planning and Teaching Science. *Journal of Research in Science Teaching, 26*(6) 467-501.

Edelson, Daniel. (2001). Learning-for-Use: A Framework for the Design of Technology-Supported Inquiry Activities. *Journal of Research in Science Teaching, 38*(3), 355-385.

Friedrichsen, P.J., Abell, S.K., Pareja, E.M, Brown, P.L, Lankford, D.M., & Volkmann, M.J. (2009). Does Teaching Experience Matter? Examining Biology Teachers' Prior Knowledge for Teaching in an Alternative Certification Program. *Journal of Research in Science Teaching, 46*(4), 357-383.

Hayes, Michael. (2002). Elementary Preservice Teachers' Struggles to Define Inquiry-based Science Teaching. *Journal of Science Teacher Education*, 13(2), 147-165.

Jacobs, C., Martin, S., & Otieno, T. (2008). A Science Lesson Plan Analysis Instrument for Formative and Summative Program Evaluation of a Teacher Education Program. *Science Education, 92*, 1096-126.

Justi, R. & Gilbert, J. (2002). Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science. *International Journal of Science Education*, 24(12), 1273–1292.

Merriam, S. B. (1988). *Case Study Research in Education: A Qualitative Approach.* San Francisco, CA: Jossey-Bass.

National Research Council. (2000). *Inquiry and the National Science education Standards: A Guide for Teaching and Learning.* Washington, DC: National Academy Press.

Peterson, P.L., Marx, R.W., & Clark, C.M. (1978). Teacher Planning, Teacher Behavior, and Student Achievement. *American Educational Research Journal, 15*(3), 417-432.

Sanchez, G. & Valcarcel, M.V. (1999). Science Teachers' Views and Practices in Planning for Teaching. *Journal of Research in Science Teaching, 36*(4), 493-513.

Schwarz, C. & Gwekwerere, Y. (2007). Using a Guided Inquiry and Modeling Instructional Framework (EIMA) to Support Preservice K-8 Science Teaching. *Science Education*, 91,158 – 186.

Shavelson, R. J. (1987). Planning. In Dunkin, M. (ed). *The International Encyclopedia of Teaching and Teacher Education* (pp. 483-486). Oxford, England: Pergamon.

Wiggins, G. Y McTighe, J. (1999). *Understanding by Design.* Alexandria, VA: Association for Supervision and Curriculum Development.

Windschitl, Mark (2004). Folk Theories of ''Inquiry:'' How Preservice Teachers Reproduce the Discourse and Practices of an Atheoretical Scientific Method. *Journal of Research in Science Teaching*, 41(5), 481–512.

Windschitl, M., Thompson, J., & Braaten, M. (2008). How Novice Teachers Appropriate Epistemic Discourses Around Model-Based Inquiry for Use in Classrooms. *Cognition and Instruction, 26*, 310-378.

Wolcott (1994). *Transforming Qualitative Data: Description, analysis, and interpretation.* Thousand Oaks, CA: Sage.

# Effects of On-line Collaborative Argumentation Processes on Justifications

Jingyan Lu, The University of Hong Kong, Pokfulam Road. HK, Jingyan@hkucc.hku.hk
Ming Ming Chiu, SUNY – Buffalo, mingmingchiu@gmail.com
Nancy Law, The University of Hong Kong, Pokfulam Road. HK, nlaw@hku.hk

**Abstract**: Justifications (through evidence or explanations) are central to productive argumentation. This study examines how the participant structures and discourse moves of students engaged in collaborative learning affect their justifications. Forty students working on Knowledge Forum, an on-line collaborative learning environment, posted 136 messages, which were coded and analyzed with an ordered logit, vector autoregression, system of equations. When students disagreed or made claims, they were more likely to use evidence. After a student made an alternative claim, the next student posting a message was less likely to use evidence. When students made claims, disagreed, disagreed with other's justifications, or read more notes, they were more likely to use explanations. Boys made more claims than girls did, but otherwise, they did not differ significantly in their likelihood of using justification. Together, these results suggest that participant structures and discourse moves are linked to justifications.

## Introduction

During argumentation, students can construct knowledge through their use of justifications to support their claims and evaluations (Lipponen, 2000; Saab, van Joolingen, & van Hout-Wolters, 2005; van Amelsvoort, Andriessen, & Kanselaar, 2007). The quality of their justifications depends on their evidence, their explanations (Clark & Sampson, 2008; Hakkarainen, 2003; Weinberger & Fischer, 2006), and the complexity of their reasoning (Hmelo-Silver & Barrows, 2008). Since justification is a critical activity while constructing knowledge, the way in which individual learners use justifications to construct claims is an important issue (Kuhn, 2001). This is especially true in online environments (Weinberger & Fischer, 2006) which can facilitate the construction of justified claims by providing learners with more time to formulate arguments (Marttunen & Laurinen, 2001); in contrast, learners rarely constructed justified claims on their own in individual argumentation due to time limitations (Kuhn, 1991).

This study focuses on messages posted on an electronic forum by students in a secondary school geography class. The students were discussing possible solutions to problems associated with global warming. The study investigated how the discourse moves of the students and the resulting participant structures affected the justifications they made during the online discussion.

## Theoretical framework

### Collaborative Argumentation and Knowledge Construction

Traditional theories of argumentation have often focused exclusively on the various stages of argumentation (Kuhn, 2001; Toulmin, 1958). Overly restrictive, they regard arguments as essentially a means to winning, overlooking its highly interactive nature. When students participate in collaborative learning, however, they typically argue to learn rather than to win. More recent approaches to argumentation have introduced the idea of *interactive argumentation*, which is defined as a social and collaborative process directed at articulating informal reasoning (Perkins, Farady, & Bushey, 1991) and constructing and advancing knowledge (Duschl & Osborne, 2002), rather than only justifying or refuting particular points of view (Van Eemeren, Grootendorst, & Snoeck Henkemans, 2002)

As a collaborative learning format, knowledge construction has shown that learners can engage in diverse forms of interactive argumentation to advance shared understandings (van Boxtel, van der Linden, & Kanselaar, 2000). The activities involved in constructing knowledge include sharing, shaping, modifying, restructuring and abandoning knowledge; all of which are both cognitive and socio-cultural. Thus, learners need environments in which they can discuss, share, evaluate, justify, and debate.

Researchers have found that learners use evidence and explanation during argumentation differently (Brem & Rips, 2000; Kuhn, 2001). For instance, Kuhn (2001) found that young children failed to differentiate between evidence and explanations, but their ability to do so increased as they grew older. Brem and Rips (Brem & Rips, 2000) found that learners could differentiate between explanations and evidence but still preferred explanations over evidence due to social and pragmatic factors. What determines the preference for explanation or evidence during argumentation? One view is that it is determined by "personal characteristics, or a person's epistemological beliefs concerning the way knowledge may be justified, and the different roles of evidence and theoretical explanation in the construction of knowledge" (Glassner, Weinstoc, & Neuman, 2005, p. 107). Moreover, the availability and strength of evidence also determines how students use it to justify arguments (Brem & Rips, 2000; Kuhn, 2001). During argumentation, if students are provided with evidence, they tend to use it. Otherwise, they tend to use explanations (Brem & Rips, 2000).

Although many studies have shown that giving evidence and explanations to justify arguments is involved in constructing knowledge, most have focused on individual rather than collaborative argumentation. Do discourse moves, such as questions, claims, and evaluations influence types of justification? Do the characteristics of learners influence the types of justification they use? The next section focuses on discourse perspectives on interactive argumentation and their relation to discourse quality.

## Discourse Moves, Participant Structures and Justifications

To examine how collaborative construction of knowledge occurs through constructive argumentation, consider two central elements: discourse moves and participant structure (Hmelo-Silver & Barrows, 2008).

Discourse moves are used here as units of analysis for investigating the socio-cultural contexts in which the cognitive processes of argumentation, such as justifications, are embedded. Discourse moves, whether written or spoken, typically involve one participant addressing one or more participants and are characterized by their communicative functions (Tapper, 1996). For instance, Hmelo-Silver and Barrows (2008) identify questions, statements, and regulations as three important discourse moves in knowledge building discourse. Thus, the discourse moves that online learners use to justify arguments often elicit further discourse moves by other students.

Question types are related to the use of evidence and explanations during collaborative knowledge construction. The presence and nature of questions in the online discourse have been used as indicators for the level of knowledge construction. Hakkarainen (1998) classified questions into factual-oriented and explanation-oriented questions, specifying that the former should yield evidence, while the latter should yield explanations (Kuhn & Pearsall, 2000). Making claims and evaluating different points of view are important discourse moves in argumentative discourse and are essential in the productive construction of knowledge. Participants make claims and evaluations to express and examine diverse perspectives, and to negotiate shared understandings in light of prior knowledge and new information (Andriessen, Baker, & Suthers, 2003; Teasley & Roschelle, 1993; Veerman, 2003).

Evaluations include full agreements, full disagreements and partial disagreements. To minimize conflicts, students typically express disagreements diplomatically through polite disagreements (Holtgraves, 1997) or unstated partial agreements and partial disagreements (Pomerantz, 1984). Based on research on face-to-face conversations, Pomerantz (1984) found that very often, agreements are preferred during conversation. Consequently, respondents usually preface a disagreement with an agreement ("Yes, … but…"). This type of partial agree/disagreement organization is often used to disagree. We will explore how evaluations affect justification in online discourse in this study.

Few studies have examined the relationship between discourse moves and types of justification. A preliminary descriptive study showed that in a collaborative learning environment, learners tend to present evidence and explanations when responding to disagreements while they tend to present more explanations when supporting claims (Clark & Sampson, 2008). However, researchers have not systematically tested explanatory models of how discourse moves or participant structures predict subsequent justifications (explanation or evidence).

The rise of collaborative learning has deepened awareness that the cognitive processes of learning are rooted in and grow out of socio-cultural processes. These include complex structures of participation by which learners constitute and are in turn constituted by communities of practice (Lave & Wenger, 1991). The notion of participant structure was originally characterized as the ways in which teachers arrange "verbal interactions with students, for communicating different types of educational material, and for providing variation in the presentation of the same material" (Phillips, 1972, p. 377).

2

The definition of participant structure focuses mainly on students' involvement or engagement, which was usually qualitatively defined or described. Participant structures in computer-supported collaborative knowledge building discourse can be represented by means of social network analysis (SNA) which can measure both reading and writing behavior (Zhang, Scardamalia, Reeve, & Messina, in press). SNA represents social networks in the form of directed graphs composed of nodes representing people, and links representing people reading or responding to the messages of other participants. Online participants may write messages that build on the messages of other participants or read messages of other participants. The positions of participants social network are based on their connections to other network participants. Researchers have also suggested that the use of evidence or explanations are also affected by learning contexts and learner characteristics (Brem & Rips, 2000). We will also explore them in this study.

This study explores the use of discourse moves and participant structures to predict two kinds of justification (giving evidence and explanations) in online argumentative discussions. Specifically, it focuses on the following four questions: 1. Are general features, such as participants' gender, age, and topics related to types of justification? 2. Are participant structures related to types of justification? Specifically, do participants with different importance to the network tend to produce different justifications? 3. Are discourse moves related to types of justification? E.g., will participants generally use more explanations when making claims or giving evaluation? Do different forms of disagreement involve different types of justification? 4. Can discourse moves predict types of justification of subsequent notes? E.g., can questions, claims and disagreements in earlier messages be used to predict types of justification in later messages?

## Methodology

### Data Sources

This study analyzes the online discourse of students in a secondary school in Hong Kong. For one semester, 40 Form Three students in the humanities course used Knowledge Forum (KF$^{TM}$) (Scardamalis, 2004), an online learning environment to support discussions. The teacher had students discuss energy problems from a number of perspectives. Two topics that had provoked sustained discussion were selected for study. One topic, "The Energy Crisis," elicited 82 postings and the other topic, "There are More Advantages than Disadvantages for China Developing Nuclear Energy," garnered 54 postings.

### Coding Schema

Utterances were coded separately for discourse moves and types of justifications. Difference levels of coding are applied for discourse moves. First, utterances are coded as claims, evaluations, questions, and information. Claims, evaluations, and information are exclusive codes. Questions are coded separately. Thus, an utterance coded as a claim, evaluation, or information may also be coded as a question. Claims may be new claims or alternative claims. Evaluation may be agreement, disagreement, agreement plus justification or disagreement plus justification. Questions may be fact-oriented and explanation-oriented.

This study examined quality of argumentation by focusing on types of justification: evidence and explanation in order to identify the extent to which students use them as warrants for claims or evaluations on others' message. Explanations involve reasoning in justification but have no empirical data, personal experience, or references to an authority, e.g., "I don't agree with your opinion because I think wind and hydro power are more environmental than coal and oil resources". Evidence involves empirical data, personal experience, or references to an authority, e.g., "I don't agree with you that there is no danger of nuclear power. The Chernobyl Nuclear Power explosion at Ukrainian in 1989 is an example". Notes that only have evidence are assigned 1 for evidence, and 0 for explanation. Notes that only have explanation are assigned 0 for evidence, and 1 for explanation. Notes that include both evidence and explanation are assigned 1s for both evidence and explanation, and notes with neither evidence nor explanation are assigned 0s for both evidence and explanation. Coding was first done by the first author and then recoded blindly by a research assistant. The inter-rater reliability, measured via agreement percentage ranged from a low of 79% (claim) to a high of 94% (question).

Various indicators were used to examine the participant structures of social networks (centrality, density) and the positions of members (frequent responders with high outdegrees vs. popular members with high indegrees, betweenness) in networks. Betweenness measured the extent of brokers' importance between two actors connected with him. Participants with high betweenness have more control over conversations.

## Analysis

Statistical analyses of group processes at the speaker turn level must overcome three difficulties. First, the outcome variable is discrete, not continuous. Second, events are often similar to recent events in time-series data (serial correlation). Third, modeling justifications requires modeling multiple outcome variables (explanation and evidence). We addressed these difficulties by using an ordered logit, vector autoregression (VAR, Kennedy, 2004), seemingly unrelated regression (SUR) systems of equations (Goldstein, 1995). We entered the variables according to time constraints, expected causal relationships, and likely importance.

$$\pi_{iy} = P(Y_{iy} = 1) = 1 / \{1 + \exp[-(\beta_{0y} + e_{iy})]\} \quad (1)$$

$\beta_{0y}$ are the grand mean intercepts $Y_{iy}$, of speaker turn $i$ for each outcome variable $y$ (Evidence and Explanation). The residuals are $e_{iy}$.

$$\pi_{iy} = 1 / \{1 + \exp[-(\beta_{0y} + e_{iy} + \beta_{vy}\mathbf{V}_{iy} + \beta_{wy}\mathbf{W}_{iy} + \beta_{xy}\mathbf{X}_{iy} + \beta_{zy}\mathbf{Z}_{iy})]\} \quad (2)$$

First, we entered a vector of $v$ speaker background variables: topic, age, gender, total number of messages, SNA aspects, such as degreeness, closeness, and betweenness of reading and building on others' notes ($\mathbf{V}$). A nested hypothesis test ($\chi^2$ log likelihood) indicated whether each set of explanatory variables was significant (Kennedy, 2004). Non-significant variables were removed. Then, we entered a vector of turn property variables: disagree, polite disagree, disagree against justification, a claim, extended claim, question, interpret question ($\mathbf{W}$). Next, we tested for the property of the previous turns. We entered lag 1 – lag 4 variables of the above turn property variables ($\mathbf{X}$). To test for moderation effects, we added interactions among significant variables ($\mathbf{Z}$).
An alpha level of .05 was used.

## Results

### Test and Summary Statistics

Of the 136 messages, 134 were coded using our coding scheme. Of the two remaining notes, one was blank, which we assumed was a system error. The other was an off-task comment and since it was the only one in the data set, we ignored it. Among the 134 notes, three were composed by the teacher, and 115 notes were coded as evaluations or claims. 16% of them were coded as having neither evidence nor explanation, 40% as having explanations only, 17% as having evidence only, and 27% as having both. 51% of the messages were posed by girls. Disagreements occurred in 25% of the messages. Also, justifications occurred more often during disagreements than during agreements. During both agreements and disagreements, explanations occurred more often than evidence use. People also used justifications more when making a claim than when evaluating a claim. Please refer Figure 1 about the effects of discourse moves and participant structure variables on evidence and explanation.

### Explanatory model

#### Predict Evidence

Boys were 33% more likely to express their ideas with evidence than girls. Girls were 13% less likely to make claims (Sobel $z = -1.97$). Taking into account the likelihood of making new claims, girls' and boys' use of evidence do not differ significantly.

When learners disagreed, they were more likely to invite evidence. A disagreement in the current turn was 39% more likely to have evidence. When learners made a claim, they preferred to use evidence. Controlling for claims in current turns, the gender effect on evidence was reduced by 30%.

Alternative claims in the previous turn yielded 8% less evidence in the current turn. When a previous learner Made an alternative claim, the next learner was 41% less likely to use evidence. The gender effect on evidence no longer significantly controlled for alternative claims in previous turns. Other variables were not significant (e.g., discussion topic).

#### Predict Explanation

Those who read more notes by other people were more likely to use explanations when expressing their ideas.

4

Disagreements in current turns were 78% more likely to have explanation. Current turns' disagreements against justification were 78% more likely to have explanation. Moreover, claims in current turns were 61% more likely to have explanation. Other variables were not significant (e.g., discussion topic).

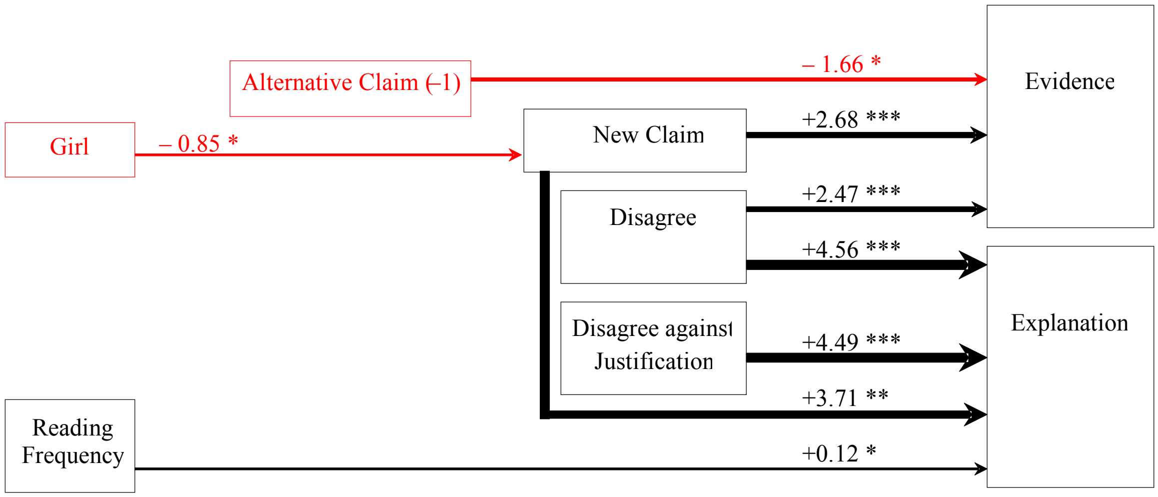



Figure 1: Path diagram of the final multivariate logit regression model predicting Evidence and Explanation. Black lines and boxes indicate positive links. Red lines and boxes indicate negative links. Thicker lines indicate larger links.

## Discussion

This study found that gender, disagreements, claims, and alternative claims can predict the use of evidence. Claims, disagreements, disagreements against justification, and students' reading frequencies are linked to the use of explanations. Messages with alternative claims yielded subsequent notes with fewer explanations.

Boys tended to use more evidence than girls in justification because they made more new claims than girls did. Taking into account the number of claims, the likelihood of evidence use did not differ significantly among boys and girls. Previous studies found that males engaged in more and longer online discussions than females, and that males tended to engage in greater levels of social exchange than females (Barrett & Lally, 1999). Females were also found to make fewer qualified arguments than males (Meyers, Brashers, Winston, & Grob, 1997). However, past studies of gender differences did not differentiate evidence vs. explanations in argumentation. Girls preferred to evaluate the opinions of others or to add their own opinions to new claims. Encouraging girls to make new claims or to state new opinions might lead to their greater use of evidence.

Participants who read more messages by others were more likely to produce explanations, possibly because they acquired more information from others to use for their explanations. In addition, reading messages from others might also help them recognize the potent impact of justifications (e.g., to persuade others to agree, e.g. Chiu, & Khoo, 2003). In contrast, participant centrality (brokerage positions) did not correlate with types of justification. The fact that brokerage positions were not related to disourse moves may indicate that such positions in online social networks do not function as they do in face-to-face social networks. Because learners in higher brokerage positions neither control nor have more access to resources, they did not perform differently from learners in non-brokerage positions.

When people make claims, they typically provide evidence and explanations to support them (Figure 3). For instance, when learners make claims, they might anticipate disagreements and provide justifications preemptively, and thereby be more persuasive (Chiu & Khoo, 2003). Disagreements are also often accompanied by evidence and explanations (especially compared with agreements) because when learners challenge each others' ideas, they are more likely to provide evidence or explanations for persuasion than when they agree with them. This finding is consistent with Clark and Sampson's (2008) study of online discussions of science issues. Furthermore, when learners challenge the validity of evidence or explanations offered for claims or evaluations, they are much more likely to use explanations than when they challenge theses (Clark & Sampson, 2008; Erduran, Simon, & Osborne, 2004). This differs from disagreeing with theses that do not attack the evidence or explanations. Learners

5

use their own explanations to undermine the beliefs of other and to justify their own opinions. The preference for explanations over evidence is consistent with previous claims that learners tend to use explanations in arguments (Kuhn, 2001) and is sensitive to goals(Glassner, et al., 2005; Schwarz & Glassner, 2003)which in this case involves disagreeing with something rather than proving it.

Alternative claims provide new interpretations of problematic phenomena. They are claims that differ from earlier claims but do not disagree with a specific aspect of the earlier claim. Unlike notes disagreeing with justifications or with claims, notes containing alternative claims do not contain suchexplicit discrepancies. The example in the results section shows that after student B's first alternative claim, student C provided a second alternative claim without evidence. It could be that people do not recognize the conflict between original and alternative claims, and without the conflict, people are less likely to use evidence after alternative claims. It could also be that because there was no strong connection between the alternative claim and the preceding message, student C was unable to follow the argument and thus was unable to provide a simple agreement or another alternative claim.

This study has theoretical, methodological and practical implications for researchers and practitioners. Theoretically, it suggests that discourse moves and participant structure might influence the use of justifications. It provides evidence that cognitive and social communicative processes are closely related and integrated in the online discourse activities. Methodologically, this study used quantitative methods to analyze messages in an online forum, traditionally viewed as qualitative discourse data. It usedordered logit, vector autoregression, systems of equations to analyze relationships among the multiple dimensions of online discourse thus modeling the discrete outcome variables and the relationships among the messages. More practically, the findings provide teachers with information for understanding and managing online discussions. Teachers usually have difficulty examining students' online discussions in detail and providing them with useful feedback. If these results are validated in future studies, then teachers can encourage students to read each others' notes to collect information and develop their skills in providing justifications.

6

## Reference

Andriessen, J., Baker, M. J., & Suthers, D. (2003). Argumentation, computer support, and the educational context of confronting cognition. In J. Andriessen, M. J. Baker & D. Suthers (Eds.), *Arguing to learn: confronting cognitions in Computer-Supported Collaborative Learning environment* (pp. 1-25). Dordrecht, Netherlands: Kluwer Academic Publishers.
Barrett, E., & Lally, V. (1999). Gender differences in an on-line learning environment. *Journal of Computer Assisted Learning, 15*(1), 48-60.
Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*(4), 573-604.
Chiu, M. M., & Khoo, L. (2003). Rudeness and status effects during group problem solving: Do they bias evaluations and reduce the likelihood of correct solutions? *Journal of Educational Psychology, 95*(3), 506-523.
Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching, 45*(3), 293-321.
Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education, 38*, 39-72.
Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education, 88*(6), 915-933.
Glassner, A., Weinstoc, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology, 75*, 105-118.
Hakkarainen, K. (1998). Epistemology of scientific inquiry and computer-supported collaborative learning. Unpublished Unpublished doctoral dissertation, University of Toronto.
Hakkarainen, K. (2003). Emergence of progressive-inquiry culture in computer-supported collaborative learning. *Learning Environments Research, 6*(2), 199-220.
Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction, 26*(1), 48 - 94.
Holtgraves, T. (1997). YES, BUT … Positive politeness in conversation arguments. *Journal of Language and Social Psychology, 16*, 222-239.
Kuhn, D. (1991). *The skills of argument*. Cambridge: University Press.
Kuhn, D. (2001). How do people know? *Psychological Science, 12*(1), 1-8.
Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development, 1*, 113-127.
Lave, J., & Wenger, E. (Eds.). (1991). *Situated learning: legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
Lipponen, L. (2000). Towards knowledge building: From facts to explanations in primary students' computer mediated discourse. *Learning Environments Research, 3*, 179-199.
Marttunen, M., & Laurinen, L. (2001). Learning of argumentation skills in networked and face-to-face environments. *Instructional Science, 29*(2), 127-153.
Meyers, R. A., Brashers, D. E., Winston, L., & Grob, L. (1997). Sex differences and group argument: A theoretical framework and empirical investigation. *Communication Studies, 48*(1), 19-41.
Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. Voss, D. Perkins & J. Segal (Eds.), *Informal reasoning and education* (pp. 83-105). Hillsdale, NJ: Erlbaum.
Phillips, S. (1972). Paaicipant structures and communicative competence: Warm Springs children in community and classroom. In C. B. Cazden, V. P. John & D. Hymes (Eds.), *Functions of language in the classroom* (pp. 370-394). New York: Teachers College Press.
Pomerantz, A. (1984). Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes In M. J. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis*. Cambridge: Cambridge University Press.
Saab, N., van Joolingen, W. R., & van Hout-Wolters, B. H. (2005). Communication in collaborative discovery learning. *British Journal of Educational Psychology 75*, 603-621.
Scardamalis, M. (2004). CSILE/Knowledge Forum. In A. Kovalchick & K. Dawson (Eds.), *Education and technology: An encyclopedia* (pp. 183-192). Santa Barbara, CA: ABX-CLIO Inc.
Schwarz, B. B., & Glassner, A. (2003). The blind and the paralytic: supporting argumentation in everyday and scientific issues. In Jerry Andriessen, Michael Baker & D. D. Suthers (Eds.), *Arguing to learn: Confronting*

*cognition in computer-supported collaborative learning environments* (pp. 227-260). Dordrecht: Kluwer Academic Publishers.

Tapper, J. (1996). Exchange patterns in the oral discourse of international students in university classrooms. *Discourse Processes, 22*(1), 25-55.

Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 229-258). Hillsdale, NJ: Lawrence Erlbaum Associates.

Toulmin, S. E. (1958). *The use of argument*. Cambridge: Cambridge University Press.

van Amelsvoort, M., Andriessen, J., & Kanselaar, G. (2007). Representational tools in computer-supported collaborative argumentation-based learning: How dyads work with constructed and inspected argumentative diagrams. *Journal of the Learning Sciences, 16*(4), 485 - 521.

van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction, 10*(4), 311-330.

Van Eemeren, F., Grootendorst, R., & Snoeck Henkemans, F. (2002). *Argumentation: Analysis, evaluation, presentation*. Mahwah, NJ: Lawrence Erlbaum Associates

Veerman, A. (2003). Constructive discussions through electronic dialogue. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (Vol. 1, pp. 117-143). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (in press). Designing for collective cognitive responsibility in knowledge building communities. *Journal of the Learning Sciences*.

# Dispositions, disciplines, and marble runs: a case study of resourcefulness

Margaret Carr, University of Waikato, Private Bag 3105, Hamilton, New Zealand, margcarr@waikato.ac.nz
Jane McChesney, University of Canterbury, Private Bag 4800, Christchurch, New Zealand, jane.mcchesney@canterbury.ac.nz
Bronwen Cowie, University of Waikato, Private Bag 3105, Hamilton, New Zealand, bcowie@waikato.ac.nz
Robert Miles-Kingston and Lorraine Sands, Greerton Early Childhood Centre, Mitchell St. & Emmett St, Tauranga, New Zealand, lorraine.sands@elp.co.nz

**Abstract:** In this paper, three researchers and two teachers have zoomed in on three 'mid-level' episodes of learning in a childcare centre and analyzed them using two lenses: a dispositional lens and a disciplinary (science) practices lens. We wonder how these two perspectives could be combined, in order to provide a coherent and illuminating analysis of the learning and to have something to say about the transactional and progressive processes towards 'being a competent learner' and 'being a scientist' that might be at work here. We have found the notion of 'resourcefulness' to be a useful boundary object; it works differently in the different lenses, but has core features in common. We look back across three documented learning episodes and teacher reflections from an early childhood centre, and consider ways in which 'resourcefulness' can provide a boundary-crossing concept that has value for discussions about learning across educational sectors.

## Introduction

Traditionally, outcomes to do with attitudes towards learning, or dispositions, have been fore-grounded in early childhood curriculum discussions, while subject- or discipline-based concerns with knowledge have been fore-grounded in secondary curriculum discussions. Two of the authors of this paper are early childhood teachers, one of the university researchers has a background in early childhood teaching, and the other two have backgrounds in secondary school teaching of science and mathematics. We are all interested in zooming in on particular episodes of learning and teaching in classrooms and early childhood centres, and then zooming out to develop ideas about the ways in which learning trajectories or learning journeys might have been set in motion or strengthened, and what learning opportunities appear to be particularly affording, inviting and engaging for this educational purpose. We all enjoy seeing rich details and the broad patterns within which the details fit: hoping to become both 'tree-wise and forest-wise' (Moss, Phillips, Erickson, Floden, Lather, & Schneider, 2009 p. 504). The university researchers had two lenses on this - learning dispositions and disciplinary practices (in this case of science). This paper focuses firstly on learning dispositions and we investigate troublesome notions such as transfer and authoring. Secondly, a lens of disciplinary practices highlights relationships between children and teachers and artifacts of the environment. We were interested that the early childhood teacher authors appeared to shift easily from one lens to another and, in this paper, we are seeking some common ideas across these two lenses, for analyzing the learning in ways that enrich our ideas about learning journeys and learning pathways.

*Learning dispositions* as a concept usefully starts with Perkins, Jay & Tishman's (1993) triad of elements for a 'thinking' disposition: inclination, sensitivity to occasion, and ability. Using the triad, dispositions can be summarized as being ready (inclined) willing (sensitive to occasion) and able (having the ability or the knowledge) (Claxton & Carr, 2004). Dispositions turn abilities into action (Ritchhart, 2002): abilities become actions when the learner is generally inclined to apply them, and sees that this might be an appropriate occasion. They appear in the literature under various names: for instance, as habits of mind (Costa & Killick, 2000), mindsets (Dweck, 1999, 2006), key competencies (Rychen & Salganik, 2003), and learning power (Claxton, 2002). They have been seen as the building blocks of a learner identity:

> How shall we deal with Self? ….. I think of Self as a text about how one is situated with respect to others and towards the world – a canonical text about powers and skills and dispositions that change as one's situation changes from young to old, from one kind of setting to another (Bruner, 1996, p. 130)

Here we build on previous research that sets learning disposition into a sociocultural frame (Carr, 2001a, b; Carr, Smith, Duncan, Jones & Lee, 2009), connecting them to situation and setting: to disciplinary practices, and to resources or affordance networks.

*Disciplinary practices* are specialized "ways of representing knowledge, and ways of thinking and inquiring that come to count as knowledge" (Kelly, Luke & Green, 2008, p. viii). In science, for instance, a valued practice of the discipline is to conduct an experiment and to repeat and refine as a means of generating knowledge. Windschitl, Thompson & Braaten (2008) write about the scientific forms of inquiry associated with the 'epistemic nature of knowledge'. "Scientists create models in the form of analogies, conceptual drawings, diagrams, graphs, physical constructions, computer simulations" and they use these models for the following forms of inquiry:

> Meaningful learning in school science entails more than accumulating domain knowledge, it requires the appropriation of specialized epistemic discourses that allow students to organize, develop, and evaluate knowledge according to disciplinary standards. (Windschitl et al., 2008, p. 310)

Models are subsets of larger, more comprehensive systems of explanation (i.e. theories) that "provide crucial frames of reference" for *generating* hypotheses or working theories for *testing*, *revising*, *conjecturing* and *explaining* (p. 311-312)

Guiding our analyses is the view that a learning activity is distributed over people, places, and things: the learner as a "person-participating-in-a-practice" (Miller & Goodnow, 1995, p. 8). In order to represent the multidimensional nature of situated activity we draw on Barab and Roth's definition of an affordance network as the "collection of facts, concepts, tools, methods, practices, agendas, commitments, and even people ... that are distributed across time and space and are *viewed* as necessary for the satisfaction of particular goal sets" (2006, p. 6) (italics added). A closely related idea is the notion of a learner as an "agent-acting-with-mediational-means" (Wertsch, 1998, p. 24) where mediational means include language, signs and symbols, and other tools and artifacts. Learners act with mediating tools by "recognising and activating a tool's affordances that are suitable" for a particular purpose or context (McChesney & Cowie, 2008, p. 108). We describe and discuss an episode about Chase rolling marbles down wooden tracks as a means of exploring how learning dispositions and disciplinary practices might inform us around issues of authoritative competence (Greeno, 2006). For example, how might a perspective based on recognizing, activating, and adapting resources encapsulate mediated activity? Finally we highlight some issues around refining these concepts, discuss the notion of 'resourcefulness', and raise further questions that might be explored in different settings.

## Background

The episodes that form the data for this paper were included in the Final Report for a three-year research project on strengthening a culture of question-asking and question-posing in an action research project by the teachers and the children (Greerton Early Childhood Centre Team, Carr & Lee, 2008). This was part of a Centre of Innovation programme, funded by the New Zealand Ministry of Education. Children who attend this childcare centre are aged from infancy until the day they go to school (usually on their fifth birthday). Significant features of the centre's practice are: a culture of reflection and critique, a philosophy of shared leadership, and an environment that affords and promotes inquiry. An aspect of the latter is their use of Learning Stories as a formative assessment mediating tool or artifact. Learning Stories are characterized by digital photographs being taken, an episode of learning is described (in text, sometimes dictated by the learner), the learning is analyzed, and suggestions for further work are made. Each Learning Story is digitally stored, printed as artifacts for portfolios, re-visited with the children by the teachers, and sent home to parents and families (Carr, 2001a). Sometimes they are written to the learner, as they were in this case. The Stories here are about Chase, who, by the time he was four years old, owned a number of portfolios of learning episodes in which he was a participant, sometimes working on his own, sometimes working in a group.

The research study involved the sixteen teachers in this centre reflecting on their Learning Stories, writing alternative perspectives, and developing a framework of key features for the affordance network associated with their intention of question-asking and question-posing. The teachers dubbed this framework the 'Threads of Inquiry'. There were six: Continuity, Listening Dialogue, Growing Intelligence, Playfulness, Surprise and Uncertainty, and Real Work. The data for this paper came from three Learning Stories in the Thread of Inquiry entitled 'Continuity' and Robert's reflections on Surprise and Uncertainty. Comments from Robert and Lorraine also come from the concluding chapter of the Final Report.

## Robert's Stories

For several months Chase has been exploring the making of 'marble runs' with angled blocks (wedges), wooden tracks and marbles. Robert, one of the teachers, writes two stories about these marble runs. In the first (What it Takes to Get There) he (Robert) had introduced a 'provocation', asking Chase whether the marble would run into a wheeled cart if it was placed at the end of the run. He and Chase began to explore together, and the story continues:

> Chase, you thought this experiment would be OK so we put the cart at the end. We found it too high so first we had to build up the end of the run high enough so the marble could drop into the cart. You tried the marble down the track. Because we had built up the end it ran out of steam before it could get into the cart.

Chase then takes over, and Robert comments:

> And here is where your keenness to experiment and your extensive knowledge of marble runs showed through. As you positioned a block and tried the marble, it still did not make it. You placed a block under the lower end, raising it up, which stopped the marble even further from the cart, this did not faze you at all but with confident persistence you continued to add a block then test, then another and test again until you had success and the marble made it into the cart. I was really impressed with your willingness to meet the challenge getting the marble into the cart, presented even when it took many attempts. Your enthusiasm (for) the task showed through in the way you celebrated by excitedly jumping and laughing when the marble finally went into the cart.

In a second story (How Far Can We Go?), two months later, including photographs of a more complex marble run, Robert writes:

> When I watch you creating one of these marble machines, it is very obvious to me, by the way you test and adjust again and again that you are always asking questions of yourself and your creation and then through testing answering them. …
>
> Also it is not only myself that has noticed your talent at marble machines but also many of the other children as they are emulating your work. In particular I have noted that Dominic's marble creations have a definite Chase influence about them.

He adds a 'Teacher Reflection'; this time the comment is not directly addressed to Chase. Other teachers and Chase's family will read these stories, and re-visit them with Chase.

> It is really great to watch how Chase is building complexity into his marble runs, and his knowledge of what is needed to make the marble follow the track he has created is growing with it. The persistence that is needed to test and retest is a disposition that will always be of use (for) problem solving whatever the task. The social side of the marble tracks cannot be ignored either. The more interesting the track you build, the more others will want to try it out or make one as well, so with limited blocks and space this provides many opportunities for negotiation, compromise, and tolerance.

## Lorraine's Story

At about the same time as these stories about the marble runs, Lorraine Sands (the second teacher author of this paper) wrote a story about three children, including Chase, rolling reels (empty electrical cable reels) down a ramp handrail. She called this story 'The Experimenters Continue'. She writes:

> Walking down the ramp is often a place of congestion. Only because it's a good place to experiment with reels and rolling. I think that today as Chase and Tom rolled and chased, rolled and chased, the appeal may have been motivated by their lengthy experimentation with pipes, marble rolling tracks and now electrical reels. They (Chase and Tom) were building their knowledge base as they worked together and alongside each other, calculating speed, distance and trajectory.

Her commentary on the learning that was happening here began with the following comment:

> I think we have experts (like Chase and Tom) who have spent long hours working out their theories, and apprentices (like Dominic) who get captured by the excitement, watch intently and then start experimenting themselves. But practice is a key element to building a knowledge base that is shaped and re-shaped through experimentation, refinement and further risk-taking in related fields – hence the fluid movement between pipes, marble tracks and electrical reels.

Lorraine notes in the documentation that she asked Chase for his 'expert opinion' about why Dominic's wider reel had veered (wobbled) off course.

> He told me immediately, "Cos it's too fat out". And he was exactly right. His endless experimentation had given him the understanding that the inside width of the reel was a critical factor in the trajectory.

## Learning dispositions

The teachers here were particularly interested in learning dispositions, which had been a feature of their early childhood curriculum since 1996 (New Zealand Ministry of Education, 1996). The national early childhood curriculum includes a strands 'Exploration'. Learning outcomes (described as a combination of knowledge, skills and attitudes) in this strand include: the ability to make decisions, choose their own materials, and set their own problems; the attitude that not knowing and being uncertain are part of the process of being a good learner; an expectation that they take responsibility for their own learning; the knowledge that trying things out, exploration, and curiosity are important and valued ways of learning. At the same time this strand includes a goal that says: "Children experience an environment where they develop working theories for making sense of the natural, social, physical and material worlds" (p. 82), and the teachers were interested in this too.

In an earlier project Carr and colleagues had taken Perkins' three elements of a 'thinking' disposition (inclination, sensitivity to occasion and ability) into the 'middle' space between the individual and the context, calling on the work of Wertsch, who writes about 'living in the middle'. He has said that "a focus on the mediated action and the cultural tools employed in it makes it possible to live 'in the middle' and to address the sociocultural situatedness of action, power and authority" (Wertsch, 1998 p. 65). We described three processes of mediated action for the construction of learning dispositions, based on Perkins' triad: authoring, recognizing opportunity, and connecting the knowing (Carr et al., 2009). These ideas build on Bourdieu's notion of *habitus* (e.g. Bourdieu 1977, 1990) and owe a considerable debt to Greeno (2006; Greeno & The Middle Schools Mathematics Through Applications Project Group, 1998), Gee (2008) and Lemke (2000). In his 2006 commentary in a special issue of the Journal of the Learning Sciences on transfer of learning, Greeno concludes (p. 546) that "To act effectively in a way that counts as transfer, therefore, involves having or taking authority to go beyond what has been taught", and we have argued that these three processes can include the seeds of learning journeys over time.

*Authoring.* Greeno (2006, p. 538) describes the conditions for authoring, authoritative and accountable positioning, as aspects of interaction such as crediting individuals with ownership, initiating ideas and topics, and challenging and questioning. Robert emphasizes persistence as a learning disposition, and comments on Chase's persistence that is needed "to test and retest", and, in a comment in the first story, 'What it Takes to Get There', he says "To be able to enter into a task with such enthusiasm and willingness to experiment is a disposition that will make so many of life's challenges easier to meet". In a commentary for the research project on the first story he commented on his observation that for Chase "a setback was just motivation to research new ideas". He interprets Chase's work as "asking questions of yourself". In a commentary for the research project on the What it Takes to Get There episode, Robert comments on what kind of an affordance is a provoking question, and whether in this case it had re-positioned Chase's authority.

> When I first looked at this, I really had to question myself. Did my posing a question about the possibility of the marble making (it to) the cart, actually support Chase in 'question-asking' when it was myself not Chase who posed the question? …. I was just thinking about how glad I was that I held my tongue while I was working with Chase as he was trying to achieve the goal of getting the marble down the marble run and into (the trailer). I found this particularly hard when Chase started to build the run up in the wrong direction".

In a commentary for the research project on the second episode, How Far Can We Go, he said: "There is a definite culture of metered intervention within the (childcare) centre. By this I mean that the teaching team here are very conscious of when and by how much they project themselves into the children's learning. They may choose to set a provocation, or work as a role model, or mentor".

*Recognizing opportunity.* After Robert had increased the challenge in this activity he reminded Chase that "You thought this experiment would be OK", and he connected the growing complexity of the marble runs with Chase's growing 'knowledge of what is needed': "His knowledge of what is needed to make the marble follow the track he has created is growing with it". He points out that it is not only the teacher "that has noticed your talent at marble machines but also many of the other children as they are emulating your work". Lorraine's story revealed that it was probably the earlier experience by Chase and Tom with water flows (pipes) and marble runs that meant that they recognized the opportunity of combining some other resources - the electrical cable reels and the ramp hand rails - for more investigations of rolling and slopes. They were re-cognizing the resources. We have described this therefore as an example of 'resourcefulness'. This depended on a range of resources being available, and a 'permeable' curriculum that gave permission for unusual and innovative combinations and uses.

*Connected knowing.* Both teachers commented on the months of practising and experimenting, the frequent episodes of exploring trajectories of rolling and flowing. They connected up the children's funds of knowledge, developed over months of practice with running water down pipes in the sandpit, and rolling marbles down marble-runs. Lorraine specifically connects the learning over time, as does Chase's portfolio of Learning Stories.

## Disciplinary practices of science

The construction of models, like marble runs, can be starting points for scientific forms of inquiry and investigation. Windschtl, Thompson & Braaten (2008) comment that "Although different domains in science have their own fundamental questions, methods, and standards for 'what counts' as evidence, they are all engaged in the same core epistemological pursuit – the development of coherent and comprehensive explanations through the testing of models" (p. 313)

In the above episodes, Chase has been developing marble runs as an interesting project, with a purpose (sending the marble into a container along a more and more complex route). From a scientist's eye, these marble runs are also models that afford scientific forms of inquiry and investigation. Chase adjusts the position of the wooden tracks by placing wooden blocks under them. This adjustment process appears to be a repeated sequence of adding more blocks (thereby adjusting the tracks) and testing with a marble run. Robert, the teacher, identifies the activity as an experiment, and acknowledges Chase's fund of knowledge about marble runs (González, Moll & Amanti, 2005). Here is Chase as scientist, engaged in a disciplinary practice of experimenting. He constructs a ramp and tries it out with a marble run. If unsuccessful, he changes parameters such as the relative heights of the ends of the tracks, or the orientation where the tracks meet, and then tries another marble run. The iterative process of adjusting and testing is at the heart of a valued scientific practice of experimenting. This particular experiment also has a desired end goal. Success is when the marble rolls down the track and drops into the wheeled cart placed at the end of the run. It is a 'eureka' moment for Chase who jumps and laughs when his goal is reached, and we would claim this excitement is also part of the human activity of being a scientist, albeit not often acknowledged in popular images of scientists.

Nested within this practice of experimenting are other aspects of scientific activity. Adjusting one end of the track more than the other changes the angle of inclination so that the track becomes a sloping ramp. It is not always a simple process of just adding blocks to raise one end of the track; depending on the conditions of the context each end of the wooden tracks may need to be raised or lowered. Chase adjusts the ramp by adding and removing different numbers of blocks at either end of the tracks. The process of adding blocks in order to adjust the slope of the ramp also provides visual information for Chase. We claim that this process of adjustment or refinement is a disciplined practice of noticing, and noticing has focus and intention. Intentional noticing is described by Mason as making "a distinction to create foreground and background, to distinguish some 'thing' from its surroundings" (2002, p. 33). We infer that Chase notices aspects of the physical artifacts of the context; the wooden tracks, the marble, the blocks, and even the wheeled cart that was to receive the marble, all have observable material attributes. These attributes are the width and length of the wooden tracks, the design of the sides of the tracks, the roughness of the wooden track, and the size and smoothness of the marble.

Associated with the physical arrangement of the tracks as an inclined plane are the aspects of motion of the marble. Chase has already noticed that the speed of the marble is an affordance of the track-and-marble system because in one of his early attempts, the marble "ran out of steam before it could get into the cart". Robert also comments in the second story:

> This latest design changes direction not once but two or three times; it will often go through tunnels and has a collection point that catches the marble at the end. To do this you explore the concepts of force, acceleration, deflection, trajectory and motion. You have worked out that the marble does not have to go fast and in fact this latest version pictured works best if the marble is rolling down just fast enough so that it will drop from one ramp to the next without overshooting. What I also find interesting about your marble runs, Chase, is that they not only function very well, but are also pleasing to look at. You seem to have a real sense for what is aesthetically pleasing.

There is more than noticing within Chase's activity. He is *recognizing which attributes are important for his experiment, such as the properties of the artifacts*. Properties of tools are affordances and according to Pea (1993), an affordance refers to "the perceived and actual properties of a thing, primarily those functional properties that determine just how the thing could possibly be used" (p. 51). The height and width of the blocks afford different aspects of being a pillar or prop for the tracks, including any limits on the placement and number of blocks used before the pillar becomes unstable. The wooden tracks afford a sloping ramp where the tracks can direct the path of the marble, and the properties of the surfaces, wood and glass, affords motion. Where each wooden track meets both affords and constrains the ramp of aligned wooden tracks. These are affordances of the tools at hand for Chase and by activating the affordances he adjusts and re-adjusts the tools until the marble run is successful. This is a reflexive process where he attends to how the blocks can raise or lower the tracks, then shifts attention to the slope of the tracks, and then back to adjusting the heights of the tracks. When he has constructed a track with the desired slope, then he shifts attention to the trajectory of the marble.

Here Chase is engaged with affordances of the track-and-marble system that are related to phenomena that we identify as aspects of physics. He is noticing, recognizing and activating affordances of the tools or resources that are suitable for his intention, and for his pursuit of scientific forms of inquiry. Recognizing the affordances of a tool plays an important role for anyone adapting the design of a tool for a particular purpose or context, as required in scientific and technological practice. It is a key to innovation and improvisation. "'Being attuned to and activating affordances' can be rephrased as 'resourcefulness', that is, noticing, recognising, recruiting, and adapting the resources of a tool for particular contexts and purposes" (McChesney & Cowie, 2008, p. 109). Particular properties or affordances of tools provide resources, where a tool is activated or re-sourced for further activity.

## Concluding comment

Episodes like these may be part of a documented journey about becoming a competent learner, and recognizing what it is to be one. They are also part of a documented journey about becoming a scientist, and recognizing what it is to be one. The marble runs discussed in this paper might be described instances of 'everyday' science, set in a context of an early childhood centre. The documentation process of Learning Stories, however, is an opportunity for the 'scientist's' view to be recorded in text and photographs. The introduction this new artifact into the environment provides a resource that will be revisited by the learner and others, be read back to the learner, and be a topic of conversations more likely to include a scientific discourse. Commentary by Lee, Brown, Brickhouse, Lottero-Perdue, Roth & Tobin (2007) reminds us that

> The repertoires of school are often compartmentalized so that students' otherwise relevant knowledge is not invited and evoked, and the continuity over time is lost to the urgency of the banal moments of everyday life in school, (p. 335)

This did not appear to be so here. The repertoires of this early childhood centre provided opportunities for inviting, evoking, and provoking spaces for the learners' relevant knowledge and experience, foregrounding continuity over time in action, dialogue, and document. As we explored different lenses on these episodes, we came to see a common, if differently conceived, element of 'resourcefulness'. This provided the common ground between the learning dispositions in situation and participation in the forms of scientific inquiry such as model-making, testing and refining working theories (aspects of the disciplinary practices of science). The commonalities were in the recognition of affordances and the re-cognizing of opportunity as a consequence of the experience that developed strategies and knowledges in the domains of disposition and science. The 'play' with pipes, marble runs, and reels provided a dispositional and scientific milieu that had the capacity to shape a learning journey. In our view the episodes illustrated that *being resourceful* is the source of improvising and reinterpreting opportunity. It is an action that is sited 'in the middle' between the cognizing and re-cognizing mind and the affording environment. And it can take learning across the boundaries of place and time.

## References

Barab, S. A., & Roth, W-M. (2006). Curriculum-based ecosystems: Supporting knowing from an ecological perspective. *Educational Researcher*, *35*(5), 3–13.

Bourdieu, P. (1977). *Outline of a theory of practice* (R. Nice, Trans.). Cambridge, UK: Cambridge University Press.

Bourdieu, P. (1990). *The logic of practice* (R. Nice, Trans.). Cambridge, UK: Polity.

Bruner, J. (1996). *The Culture of Education*. Cambridge, MA: Harvard University Press

Carr, M. (2001a). *Assessment in early childhood settings: Learning stories*. London: Paul Chapman.

Carr, M. (2001b). A sociocultural approach to learning orientation in an early childhood setting. *Qualitative Studies in Education*, *14*(4), 525–542.

Carr, M., Smith, A. B., Duncan, J., Jones, C., Lee, W. & Marshall, K. (2009) *Learning in the Making: disposition and design in early education*. Rotterdam: Sense Publishers.

Claxton, G. (2002). *Building learning power: Helping young people become better learners.* Bristol, UK: TLO Limited.

Claxton, G., & Carr, M. (2004). A framework for teaching learning: The dynamics of disposition. *Early Years*, *24*(1), 87–97.

Costa, A. L., & Killick, B. (2000). *Discovering and exploring habits of mind.* Alexandria, VA: Association for Supervision and Curriculum Development.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality and development.* Philadelphia: Psychology Press (Taylor & Francis group).

Dweck, C. S. (2006). *Mindset: The new psychology of success.* New York: Random House

Gee, J. P. (2008). A sociocultural perspective on opportunity to learn. In P. A. Moss, D. C. Pullin, J. G. Gee, E. H. Haertel, & L. Jones Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 76–108). Cambridge, UK: Cambridge University Press.

Greerton Early Childhood Centre Team, Carr, M., & Lee, W. (2008). *A Question-asking and Question-exploring Culture*. Unpublished manuscript. Hamilton, University of Waikato Wilf Malcolm Institute of Educational Research.

González, N., Moll, L. C., & Amanti, C. (2005). *Funds of knowledge: Theorizing practices in households, communities and classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates.

Greeno, J. G. (2006). Authoritative, accountable positioning and connected, general knowing: Progressive themes in understanding transfer. *The Journal of the Learning Sciences*, *15*(4), 537–547.

Greeno, J. G., & The Middle School Mathematics Through Applications Project Group. (1998). The situativity of knowing, learning and research. *American Psychologist*, *58*, 5–26.

Kelly, G. J., Luke, A., & Green, J. (2008). Introduction, What counts as knowledge in educational settings: Disciplinary knowledge, assessment, and curriculum. *Review of Research in Education*, *32*, vii-x.

Lee, Y-J, Brown, B., Brickhouse, N., Lottero-Perdue, P., Roth, W-M., & Tobin, K. (2007). Discursive constructions of identity. In W-M. Roth & K. Tobin (Eds.), *Science, learning, identity: Sociocultural and cultural-historical perspectives* (pp. 325–337). Rotterdam, The Netherlands: Sense Publishers.

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture, and Activity*, *7*(4), 273–290.

McChesney, J. & Cowie, B. (2008). Communicating, thinking and tools: Exploring two of the key competencies. *Curriculum Matters, 4*, 102-111.

Mason, J. (2002). *Researching your own practice: The discipline of noticing*. London: RoutledgeFalmer.

Miller, P. J., & Goodnow, J. J. (1995). Cultural practices: towards an integration of culture and development. In J. J. Goodnow, P. J. Miller, & F. Kessel (Eds.) *Cultural practices as contexts for development* (pp. 5-17). San Francisco: Jossey-Bass.

Moss, P. A., Phillips, D. C., Erickson, F. D., Floden, R. E., Lather, P., & Schneider, B. L. (2009). Learning from our differences: A dialogue across perspectives on quality in education research. *Educational Researcher, 38*(7), 501-517.

New Zealand Ministry of Education. (1996). *Te whäriki. He whäriki mätauranga mö ngä mokopuna o Aotearoa: Early childhood curriculum.* Wellington, NZ: Learning Media.

Pea, R. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.),*Distributed cognitions: Psychological and educational considerations* (pp. 47-87). New York: Cambridge University Press.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly*, *39*(1), 1–21.

Ritchhart, R. (2002). *Intellectual character: What it is, why it matters, and how to get it.* San Francisco: Jossey-Bass.

Rychen, D. S., & Salganik, L. H. (2003). *Key competencies for a successful life and well functioning society.* Cambridge, MA: Hogrefe & Huber.

Wertsch, J. V. (1998). *Mind as action.* Oxford, UK: Oxford University Press.

Windschitl, M., Thompson, J., & Braaten, M. (2008). How novice science teachers appropriate epistemic discourses around model-based inquiry for use in classrooms. *Cognition and Instruction, 26*(3), 310-378.

## Acknowledgments

We acknowledge funding from the New Zealand Ministry of Education for the Centres of Innovation programme.

# Reading in the Context of Online Games

Constance Steinkuehler, Catherine Compton-Lilly, & Elizabeth King
University of Wisconsin-Madison, 225 North Mills Street, Madison WI 53706
constances@gmail.com, comptonlilly@wisc.edu, emking29@gmail.com

**Abstract:** Research suggests that text is an important component of videogame culture (Gee, 2003; Leander, & Lovvorn, 2006; Steinkuehler, 2007), but we have few empirical assessments of what kinds of texts are involved or youth's reading performance on them. This paper presents a series of four studies conducted to examine: What texts are a regular part of videogame play? What is their nature, function, and quality? And what is the nature of adolescent reading performance within such contexts? Our results show that *informational* texts, comprised of 20% academic language and with an 11.8 average reading level, are the most prevalent text type used for gameplay. Reading performance on such texts is no different than on school-related texts when topic and difficulty were controlled. However, when struggling readers when allowed to *choose the topic*, they performed at "independent" level (94-97% accuracy) even on texts that were 7-8 grade levels above their head.

In their recent *Reading at Risk* report (Bradshaw & Nichols, 2004), the National Endowment for the Arts documented a pronounced decline in reading in American youth and associated it with the equally pronounced increase in the use of "electronic media" such as television, videogames, and the Internet. Such technologies, the report authors claimed, "often require no more than passive participation" and "foster shorter attention spans and accelerated gratification" (p. vii); thus, the authors identified them as one (perhaps *the*) underlying cause of this new "risk" to reading among today's digital immersed youth. Emerging research on what youth actually do while engaged in such technologies, however, reveals a rich array of digital and print literacy practices (Black & Steinkuehler, 2007). They use text messages as a way to develop and sustain peer networks (Lewis & Fabos, 2005). They create elaborate textual representations of themselves on personal websites as an expression of identity (Guzzetti, 2006). They use the given narrative structures of anime, for example, as fodder for the articulation (and distribution) of their own stories (Black, 2008). And they engage in complex constellations of reading and writing around the videogame titles they play (Steinkuehler, 2007).

The goal of this investigation was to examine this latter phenomenon – reading related to naturally occurring online videogame play. Emerging research suggests that text is an often-hidden but nonetheless important component of participation in videogame culture (Commeyras, 2009; Gee, 2003; Leander, & Lovvorn, 2006; Steinkuehler, 2006, 2007, 2008), but we have few empirical assessments of what kinds of texts are involved in such activities, their quality and characteristics, or youth's reading performance on them compared to texts in other contexts. Precisely what kinds of texts are a regular part of videogame play? What is the nature, function, and quality of such texts? And how does the reading performance of adolescents on such game-related texts compare to their performance on school-related texts? Toward answering these questions, we conducted a series of investigations designed to explore the claims popular in "games and learning" literature that videogames do indeed entail rich forms of literacy. In this paper we detail a series of four studies we conducted in order to answer the three questions articulated above: (1) a survey of the textual resources gamer players used as a regular part of their gameplay; (2) a quantitative and qualitative evaluation of the nature, difficulty, and quality of the text resources identified in study one; (3) a study comparing adolescent reading performance on appropriately leveled reading materials in the context of games versus school; and (4) a preliminary (yet suggestive) pilot study of videogame reading when difficulty level is not controlled. All four studies were conducted using the massively multiplayer online game *World of Warcraft* (WoW) given its overwhelming success on the market (Woodcock, 2009), robust online textual community, and general popularity with adolescents and young adults. In what follows, we detail the method and findings of each study.

## Study One: What Texts are a Regular Part of Gameplay?

### Data Collection

Our first step in this investigation was to find out what texts were involved regularly in gameplay. Toward this end, we interviewed 25 expert and 21 novice WoW gamers (n=46) about the textual resources they used for gameplay. Each participant was asked to list to exhaustion any and all resources they used for the game and to judge the importance of each text listed by classifying it as either a "core resource" (a text they consult regularly and consider "core" to successful gaemplay), a "frequent resource" (a text they refer to frequently but do not consider crucial) or an "infrequent resource" (a text they refer to but not with any regularity).

### Findings

Using the resources generated by participants as our basis, we then inductively developed the following five categories of text based on function:

(1) *Information resources* are online reference materials providing detailed information about the game, strategies for playing it, and the history of the game's lore. The three most common resources referenced were wowhead.com, wowwiki.com, and thotbot.dom, which are database backed websites that organize information about in-game materials and quests that are written, harvested, and then discussed, revised, and scrutinized by the players themselves. Such sites are structured as encyclopedic, dynamically generated, online reference books and function as user manuals to the game that are created and maintained by those who play.

(2) *Discussion forums,* sponsored by the game company or various public or private sub-communities within the game, host text-based, asynchronous collective problem solving about complex systems in the game – activities which evidence informal science literacy characteristics such as evidence-based argumentation and model-based reasoning (Steinkuehler & Duncan, 2009).

(3) *Group organization resources* are online tools and texts that help groups within the game (such as guilds or raid groups) organize their collaborative gameplay, providing detailed instructions for specific activities, scheduling information and sign-up tools, for example.

(4) *UI mod sites* (short for "user interface modifications," also called "add on's") are online hubs for freely downloadable, player-generated bits of software that change the user interface of the game in some way (e.g., wowcurse.com) to enhance performance or player experience. Prior case study work suggests that, for some players, the use of such tools transitions into critique, debugging and "beta testing" practices that provide an important gateway into software development practices (Steinkuehler & Johnson, 2009).

(5) Finally, *fandom media* include creative fiction and non-fiction digital media literacy artifacts such as You Tube movies, fan fiction about the game, fan artwork, and other creative "literary" resources that the player community creates around the game.

Table 1 presents the each functional category of resource by number of unique members that were referenced during the interviews, their importance rating, the average number of references per member of the category, and an average weighted score of their importance calculated as [(number of core ratings)*10 + (number of frequent ratings)*3 + (number of infrequent ratings)*1] / (total number of references)].

Table 1: Game-related texts by functional category and rating of importance to gameplay.

| **Category** | **Unique Members** | **Importance Rating** | | | | ***Average Number of References per Member*** | ***Average Weighted Importance Rating*** |
|---|---|---|---|---|---|---|---|
| | | Core | Frequent | Infrequent | Total | | |
| **Information Resources** | **31** | 52 | 55 | 77 | **185** | ***5.97*** | *4.11* |
| **Discussion Forums** | **31** | 11 | 13 | 32 | **56** | *1.81* | *3.23* |
| **Group Organization Resources** | **34** | 31 | 15 | 21 | **48** | *1.41* | ***7.8*** |
| **UI Mod Sites** | **16** | 18 | 9 | 18 | **45** | *2.81* | *5.0* |
| **Fandom Media** | **9** | 9 | 4 | 7 | **20** | *2.22* | *5.45* |

*Information resources* were by far the most common, with participants referencing unique members within this category three times more frequently than any other type of resource (roughly six citations for each member on average compared to two). In contrast, *group organization resources* were rated as more important to regular gameplay with an average weighted importance rating (7.8) followed by player production and UI-Mod resources (5.45 and 5.0 respectively), and finally information resources and discussion forums (4.11 and 3.23 respectively). Figure 2 below shows the proportional make up of the "constellation of literacy" that comprises online gameplay (Steinkuehler, 2007). Thus, information resources comprise the majority of the texts related to games but group organization resources are viewed as more important to regular gameplay. The result is a complex picture where the *main body of game-related text* consists of standard expository and procedural text with the *most valued resources* being for in-group coordination.

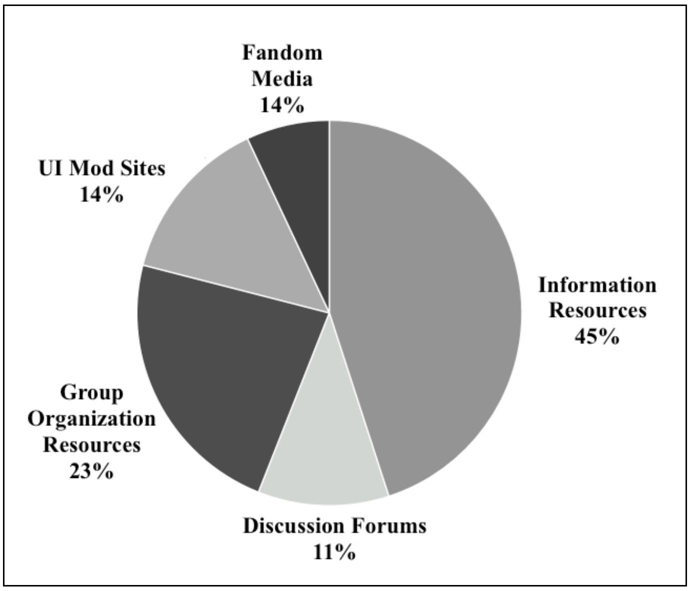



Figure 1. Proportional breakdown of the "constellation of game-related texts" by functional category.

## Study Two: What is the Nature and Quality of Game-Related Texts?

### Methods

Using the list of resources obtained in the first study, we identified the top three resources used by gamers as part of their play based on the total number of references: thotbot.com, wowwiki.com, and wowhead.com. All three similarly structured, database-backed websites function primarily as *information resources*. In order to assess the nature and quality of such texts, we randomly selected 50 pages total across the three identified resources for quantitative assessment (grade-referenced reading level, lexical analysis) and one representative page for detailed qualitative (multimodal, structural, and genre) analysis. Quantitative assessment of the 50-page random sample included statistical calculation of average reading level using three separate statistical approaches – Flesch-Kinkaid reading level (Flesch, 1948), Fry Readability (Fry, 1968), and SMOG reading level (McLaughlin, 1969) – and a lexical analysis (Laufer & Nation, 1995) to highlight the range of vocabulary used in such texts. Qualitative analysis of the single representative page included a survey of the modes of information used, the structure of the information presented, and the genres of text included.

### Findings

The average reading level of the 50-page random sample was $M$ = 10.86 ($SD$ = 2.30) on the Flesch-Kinkaid, $M$ = 12.98 ($SD$ = 2.08) on the Fry, and $M$ = 11.56 ($SD$ = 2.56) on the SMOG test. Thus, the most prevalent texts related to WoW are written at an average grade level of 11.8, placing is somewhere between *Sports Illustrated* (grade 11) and *Time Magazine* (grade 12). Although there is some variation in the reading level of game-related texts (roughly 2 grade levels), overwhelmingly such information resources require reading skills at the high school or high school graduate level.

Lexical analysis of the same 50-page sample reveals that 70% of the vocabulary used within the texts were words found within the most common 1000 words of the English language (K-1 words). A full 20% of the vocabulary used were "academic words" (Academic Word List words) – vocabulary common to school textbooks across a range of academic disciplines such as "compensation," "implication," and "obtainable". Only 4% consisted of game-related "specialized discourse" (Gee, 2003)– highly specialized terms relevant only to WoW gameplay itself such as "respec," "frost," "crit," (Off-List words). Thus, as predicted by Gee (2003) and others, game-related texts appear to be a useful bridge into academic forms of language.

Qualitative assessment of the one page representative sample nature reveals that such game-related texts are dense and complicated information sources in at least three ways. Figure 2 gives a simplified diagram articulating some of this complexity. First, symbolic information is thoroughly multimodal, including images, icons, symbols, color and font demarcations to specify information type, and hyperlinks to additional details (which display as call-out boxes on one's cursor) or link to related pages within or beyond the site itself. Second, only tacit page layout characteristics are used to cue the reader as to the source of information contained within a given page. The center of the main page features information "harvested" from the game engine automatically through site-related UI mods or "add-ons" that automatically users voluntarily use while playing. These "add-ons" collect information about the game and contribute those data back to the website database to be aggregated automatically as content within the main pages. The bottom of the page, however, features multiple tabbed sections including crucial threaded commentary and discussion of the information harvested and then re-presented in the main body of the page. In effect, the bottom half of the page provides human interpretation and analysis of the machine-generated data presented in the top half, creating a complex "split attention" effect (Mwangi & Sweller, 1998) requiring readers to synthesize information from two sources quite disparate in terms of multimodal tendencies, page layout formats, and underlying source or author.

Third and finally, such texts are particularly complicated due to their mix of genre. While the primary genres of game-related information resources are *expository* and *procedural,* each page within such sites actually features multiple forms – including transactional, narrative, and persuasive text as well as expository and procedural – with few clear demarcations among them. For example, a description of a single in-game quest very often begins with details about where the activity begins and ends (expository text), then shifts to quoted text delivered orally by the relevant "quest giver" character in the game world (narrative, first person text), followed by step-by-step instructions on how to complete the quest (procedural text), which are then quickly succeeded by player-written, collaborative arguments for why a given quest is important or not important to complete (persuasive texts), and finally, drifting conversation where players best one another in reports of how easy a given quest was overcome (transactional texts). Thus, if structural cues are one tool a reader uses to help predict and parse the content of a given text (Coiro & Dobler, 2007; Goldman & Rakestraw, 2000; Weaver & Kintsch, 1991), then multimodal, multi-authored, and multi-genre texts such as these present interesting challenges, even for strong readers.

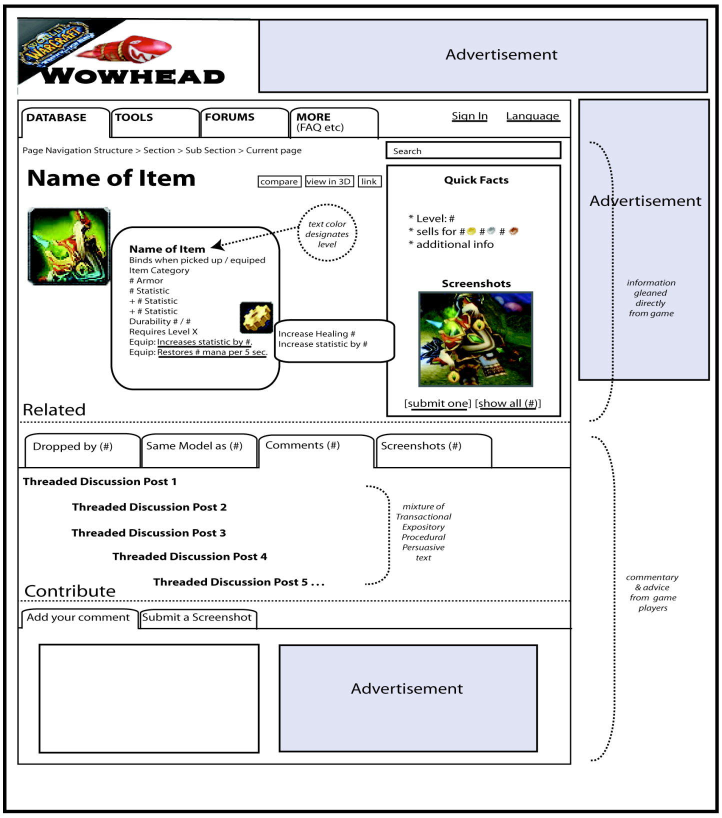



Figure 2. A schematic overview of the game-related sample page selected for qualitative analysis.

## Study Three: How Does Reading Performance Compare on Game-Related Texts versus School-Related Texts?

### Methods

Our third step in this investigation was to examine adolescent WoW players' reading performance (Clay, 1993/2002; Ekwall & Shanker, 1993; Goodman, Watson, & Burke, 2005) and comprehension strategies (Pearson & Johnson, 1978; Taylor, Graves, & van den Broek, 2000) on game-related text versus school-related text. Seventeen adolescent boys participated in the study, all of whom played WoW. Participants' reading levels were first assessed using the Qualitative Reading Inventory (QRI) (Leslie & Caldwell, 2006). They then read two texts aloud: (1) a passage from a social studies textbook, and (2) a game-related text that had been downloaded from one of the resources identified in study one. Each of the two texts were chosen by the resarchers to match the "instructional reading level" of the participant as determined by the QRI and the order of the two test texts were counterbalanced to mitigate ordering effects. The activity took about 90 minutes on average. Data included running records (Clay, 1993/2002) which noted the various reading behaviors of the students including substitutions they made in text, self-corrections, repetition of words and sentences, and attempts at word solving. Text retellings and responses to comprehension questions were used to assess reading comprehension. Finally, short interviews were conducted with each student to assess students' attitudes toward the texts and their perceived potential uses in the lives of students.

## Findings

Table 3 shows the quantitative findings from study three. All 17 adolescent boys who participated in this study read at or below grade level (n=10 and n=7, respectively). With reading selection calibrated to match each participants' reading level as determined by the QRI, there were no statistical differences between participants' accuracy or comprehension on the game related text ($M$ = 0.96, $SD$ = 0.03) versus the school related text ($M$ = 0.97, $SD$ = 0.02). Overall then, reading performance was remarkably consistent across the two texts.

Table 2. Participants' reading performance on the QRI, the game-related text, and the school related text.

| Student | Age | Grade | QRI Reading Assessment | QRI Text: Miscues | QRI Text: Compre-hension | Game Text: Level | Game Text: Miscues | School Text: Level | School Text: Miscues |
|---|---|---|---|---|---|---|---|---|---|
| Andy | 11 | 5 | 5 | 0.95 | 0.75 | 8 | 0.94 | 7 | 0.89 |
| Ben | 13 | 7 | 7–8 | 0.94 | 0.90 | 8 | 0.89 | 7 | 0.96 |
| Carl | 13 | 7 | 6 | 0.98 | 0.87 | 8 | 0.97 | 7 | 0.98 |
| Dexter | 12 | 7 | 7–8 | 0.98 | 0.80 | 8 | 0.95 | 7 | 0.98 |
| Ed | 14 | 8 | 5 | 0.95 | 1.00 | 8 | 0.92 | 7 | 0.94 |
| Fred | 14 | 8 | 6 | 0.95 | 0.87 | 8 | 0.94 | 7 | 0.97 |
| Gary | 14 | 9 | 6 | 0.99 | 0.87 | 8 | 0.99 | 7 | 0.98 |
| Harry | 14 | 9 | 7–8 | 0.97 | 1.00 | 8 | 0.95 | 7 | 0.93 |
| Isaac | 16 | 10 | 7–8 | 0.97 | 0.90 | 8 | 0.95 | 7 | 0.96 |
| Jack | 16 | 10 | HS | 0.97 | 0.90 | 10 | 0.99 | 7 | 0.99 |
| Kurt | 16 | 10 | HS | 0.98 | 0.80 | 11 | 0.99 | 12 | 0.98 |
| Lee | 16 | 10 | HS | 0.99 | 0.90 | 11 | 0.99 | 12 | 0.98 |
| Mark* | 16 | 11 | 6 | 0.98 | 0.75 | 8 | 0.97 | 7 | 0.96 |
| Ned | 16 | 11 | HS | 0.98 | 0.70 | 10 | 0.98 | 7 | 0.96 |
| Otto | 17 | 11 | HS | 0.98 | 0.80 | 11 | 0.99 | 12 | 0.98 |
| Pete* | 16 | 11 | HS | 0.98 | 0.90 | 11 | 0.99 | 12 | 0.98 |
| Quincy | 18 | 12 | HS | 0.95 | 0.80 | 11 | 0.96 | 12 | 0.98 |

We then identified two contrasting cases from within our data set (see asterisked rows in Table 3) for more detailed qualitative exploration in order to explore potential differences in participants' reading comprehension strategies and how they positioned themselves relative to the two text types. Pete (pseudonym) was in grade 11 and appropriately reading at the high school level; Mark (pseudonym) was also in grade 11 but reading at only the sixth grade level. While Pete scored 0.90 on the QRI measure of reading comprehension, Mark only scored 0.75. We transcribed and reiteratively coded each boy's retelling of the QRI text, the game text and the school text to identify similarities and differences between the two participants.

Both Pete (stronger reader) and Mark (struggling reader) used the structure of each reading to organize their retelling, but Pete adhered to school expectations when retelling school text, providing sundry details on each and given concise overviews at the beginning of each retelling episode. In contrast, Mark provided only inaccurate and sparse details on each reading and gave unclear and inconcise overviews for each text. Even when Mark (struggling reader) seems to understand the content of what was read, he had a hard time conveying it to tester. Both participants use I-statements in relation to the gaming text, but Pete (stronger reader) took on the role of “text analyst” of sorts, readily assuming an analytic position toward both texts and freely giving his opinion on the authors’ craft. In contrast, Mark (struggling reader) did not position himself as an evaluator or critic of text except in relation to his own purposes and needs for the content the text contained. For example, while discussing the school-related textbook excerpt, Pete (stronger reader) states, “I thought it was quite informative, used fairly simple and straight forward vocabulary and um, sentence structure and the like. So it was easy to understand, easy to tell it’s from a textbook.” (Pete, stronger reader) In contrast, Mark (struggling reader) discusses the purpose of the text only when directly queried, stating, “Our vocabulary…um, cause there were a lot of words in here that you had to try at” except when that assessment related to his own purposes and needs. When his own interests are evoked, his responses and evaluations become more elaborated. For example, when asked the purpose of the game-related text, he states “to explain the use of duel wielding - off-handed, one-handed, two-handed. To explain damage meters, whether DPS is better with two weapons or one two-handed weapon. To help you with the numbers I was gonna use it on my um, warrior and then my rouge to, like, go through different weapons I have in my bags and see which ones are better.” Thus, one notable difference between the two boys who contrasted in terms of reading comprehension performance was the extent to which each positioned himself as a critical consumer of the text: Mark (stronger reader) analyzed and critiqued the text’s composition and rhetoric without restriction while Pete (struggling reader) only engaged in such analysis when it was directly related to his own interests, preferences, and uses of the content of the text.

## (Pilot) Study Four: How Well Do Struggling Readers Perform on Difficult (Game-Related) Texts of Their Own Choice?

### Methods

Based on the qualitative differences found in study three – with contrasting readers (Mark and Pete) positioning themselves differently toward the text except when individual interests, preferences, and uses of the text content were evoked – we decided to conduct a follow-up study that might take individual choice (and therefore agency) better into account. In study three, we chose the text participants read and ensured that each text (school-related or games-related) was appropriately leveled to their reading ability as determined by the initial QRI. In study four, currently only in pilot phase [1], we allowed the participants to *choose the topic of the text* and *did not artificially lower its difficulty* (recall that the average reading level of game-related texts was found to be 11.8).

Three struggling readers from study three were recruited for participation. We asked each boy to specify three game-related topics he was interested in reading about. We then selected one reading for each boy that met the following criteria: (i) it was on a topic the given participant expressed interested in, (ii) it was from one of the three most frequently cited information resources identified in study one, (iii) was at least as difficult as game-related text generally (i.e., reading level of 11.8 or higher), and (iv) was at least two grade levels above the participants current reading level (as determined by the QRI in study three). Each participant answered preliminary questions assessing their attitudes toward reading generally and prior knowledge about the topic and then read the text aloud while the researcher took a running record (Clay, 1993/2002) using the same procedures as in study three. Participants were then asked to give a retelling of the text (used to assess reading comprehension) and to answer two follow-up questions assessing attitudes toward the text topic itself.

### Findings

Table 3 presents the pilot findings of this fourth and final study. All three participants were struggling readers with two participants reading three levels below their grade in school and one participant reading five levels below their grade in school. The game texts used ranged from grade level 12 in difficulty to grade level 14 (college), meaning that the boys were asked to read text that was seven to eight grade levels above their current reading level. Despite this large difference between text difficulty and current reading level, however, all three boys read at the “independent” level with an accuracy rate of 94-97%.

Table 3. Participants’ reading performance on difficult game-related text when individuals chose the topic.

| Student | Age | Grade | QRI Reading Assessment | Δ (Grade – Reading) | Game Text Level | Δ (Text Level – Reading Level) | Accuracy |
|---|---|---|---|---|---|---|---|
| Ed | 14 | 8 | 5 | -3 | 12 | + 7 | 94.0% |
| Fred | 14 | 9 | 6 | -3 | 13 | + 7 | 98.0% |
| Mark* | 16 | 11 | 6 | -5 | 14 | + 8 | 97.0% |

Of those errors that were made, the majority consisted of vocabulary words that were unfamiliar to the participants (e.g. "coup d'etat") or words they had read silently before but had never heard aloud (e.g. "subtle"). All three readers exhibited a high percentage of self-correction rates (37%, 31%, and 57% respectively), suggesting that one potential source of the performance difference between study three (assigned texts, reading difficulty controlled) and study four (texts on topics of their choice, reading level substantially higher than their current purported ability) is an increased willingness to correct errors during reading performance. In other words, in the face of challenge or difficulty, it may be that increased interest in the content of the text results in increased persistence in the face of frustration leading to higher overall achievement.

## Conclusions

While it may sometimes be tempting to see "electronic media" like videogames as something distinct from reading and in competition with it, a view expressed unequivocally within the *Reading at Risk* report (Bradshaw & Nichols, 2004), our findings suggest that games (and perhaps similar media) are not replacing text reading but rather sit in a complex and productive ecology with them. Perhaps like all new media before them, games simply highlight the enduring importance of text and text based literacy practices to everyday life. In study one, we found that, while group organization resources were considered the most important form of textual resource for regular WoW gameplay, *information resources were the most prevalent* type of text used. Closer analysis of such information resources in study two revealed that such materials read at the high school graduation level (11.8), which places them in terms of overall difficulty somewhere between *Sports Illustrated* and *Time Magazine*. This higher level of difficulty stems not from the use of highly specialized "gamer" discourse (which accounted for only 4% of the text overall) but rather from frequent use of academic vocabulary (20% of the text sample) and densely multimodal, multi-authored, and multi-genre nature. In study three, we examined adolescent reading performance on such texts versus adolescent reading performance on school textbook samples of comparable difficulty in order to ascertain whether or not there were any differences between these two contexts. Quantitatively, we found surprising consistency across both texts types in terms of accuracy and comprehension. Subsequent qualitative contrasting case analysis revealed no notable contrasts in the way one strong versus struggling reader positioned themselves toward the text and task *except* when the reader's own interests, needs, and uses for the text were evoked. Therefore, we conducted a fourth and final study (in pilot phase at the time of this writing) in which participants were able to choose the topic of the text and where we did not artificially ensure that its difficulty would be appropriate to their reading level. What we found was surprising in some ways yet predicted (Gee, 2003): All three "struggling" readers, when allowed to choose the topic, read text seven to eight grades above their reading level (as diagnosed by the QRI) with 94-97% accuracy rate – the highest level of mastery: independent.

Together, a picture emerges from these findings in which the reading activities that occur as a regular part of videogame play entail informational texts that include academic language and are structurally complex. Reading such game-related materials appears to entail the same reading performances and processes of reading activities required in classrooms. However, the fact that that activities related to them are *interest driven* appears to give them a serious edge over assigned and selected (school type tasks), potentially in particular for readers who are otherwise "struggling." This notion is not all that new, although this may be the first empirical evidence for it of this particular ilk. Scholars who study games and learning have noted that children and young adults, within their affinity spaces (Gee, 2004), often read text that is far more complicated and difficult than what they (are willing to) read in school. If such is the case, and our preliminary findings suggest it is, then we need to ask ourselves to what extent we are testing "interest" and calling it "ability." We may also need to consider whether the most powerful aspect of "games based learning" is not the complex system simulations they entail or the rules or roles or systems that constitute them but rather the fact that they are, in a way, Trojan horses for bringing interest-driven learning back into the frame of what we, as educators and learning science researchers, might think worth serious consideration.

## Endnotes

[1] At this time, we have completed only the pilot study with three participants; the full study (n=20) will be conducted in early November 2009, after submission of this paper but before the final conference event.

## References

Black, R.W. (2008). *Adolescents and Online Fan Fiction*. New York: Peter Lang.
Black, R. W. & Steinkuehler, C. (2009). Literacy in virtual worlds. In L. Christenbury, R. Bomer, & P. Smagorinsky (Eds.), *Handbook of Adolescent Literacy Research* (pp. 271-286). New York: Guilford.
Bradshaw, T., & Nichols, B. (2004). *Reading At Risk: A Survey of Literary Reading in America (Research Division Report No. 46)*. Washington, DC: National Endowment for the Arts.
Clay, M. M. (1993/2002). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.

Coiro, J. & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly, 42*(2), 214-257.

Commeyras, M. (2009). Drax's Reading in Neverwinter Nights: with a tutor as henchman, *eLearning, 6*(1), 43 - 53.

Ekwall, E. E., & Shanker, J. L. (1993). *Ekwall/Shanker reading inventory* (3rd ed.). Boston, MA: Allyn & Bacon.

Flesch, R.(1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.

Fry, E. (1968), A readability formula that saves time. *Journal of Reading, 11*(7), 265-71

Gee, J. P. (2003). *What videogames have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Gee, J.P. (2004). *Situated language and learning: a critique of traditional schooling*. New York: Routledge.

Goldman, S.R., & Rakestraw, J.A. (2000). Structural aspects of constructing meaning from text. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311-335). Mahwah, NJ: Erlbaum.

Goodman, Y., Watson, D. & Burke, C. (2005). *Reading Miscue Inventory: From Evaluation to Instruction*. Second Edition. Portsmouth, NH: Heinemann.

Guzzetti, B. J. (2006). Cybergirls: Negotiating social identities on cybersites. *eLearning, 3*(2), 158-169.

Laufer, B., & Nation, P. (1995). Vocabulary size & use: Lexical richness in L2 written productions. *Applied Linguistics 16*(3), 307-322.

Leander, K., & Lovvorn, J. (2006). Literacy networks: Following the circulation of texts and identities in the school-related and computer gaming-related literacies of one youth. *Curriculum & Instruction, 24*(3), 291–340.

Leslie, L., & Caldwell, J. (2006). *Qualitative reading inventory - 4th edition*. Reading, WA: Allyn and Bacon.

Lewis, C., & Fabos, B. (2005). Instant messaging, literacies, and social identities. *Reading Research Quarterly, 40*(4), 470-501.

McLaughlin, G.H. (1969). SMOG grading: A new readability formula. *Journal of Reading, 12*(8) 639-646.

Mwangi, W. & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition & Instruction, 16*(2), 173-199.

Pearson, P. & Johnson, D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart and Winston.

Steinkuehler, C. A. (2006). Massively multiplayer online videogaming as participation in a discourse. *Mind, Culture, & Activity, 13*(1), 38–52.

Steinkuehler, C. (2007). Massively multiplayer online gaming as a constellation of literacy practices. *eLearning*, 4(3) 297-318.

Steinkuehler, C. A. (2008). Cognition and literacy in massively multiplayer online games. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of Research on New Literacies* (pp. 611–634). Mahwah, NJ: Erlbaum.

Steinkuehler, C. & Duncan, S. (2009). Informal scientific reasoning in online virtual worlds. *Journal of Science Education & Technology*. DOI: 10.1007/s10956-008-9120-8.

Steinkuehler, C. & Johnson, B. Z. (2009). Computational literacy in online games: The social life of a mod. *The International Journal of Gaming and Computer Mediated Simulations, 1*(1), 53-65.

Taylor, B., Graves, M. & van den Broek, P. (2000). *Reading for meaning: Fostering comprehension in the middle grades*. New York: Teachers College Press.

Weaver, C.A., III, & Kintsch, W. (1991). Expository text. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 230-244). White Plains, NY: Longman.

Woodcock, B.S. (2009). *An analysis of MMOG subscription growth 23.0*. Retrieved April 21, 2009 from http://www.mmogchart.com

## Acknowledgments

This work was made possible by a grant from the MacArthur Foundation. We would like to thank Seann Dikkers and Sean Duncan for their assistance in data collection in Study One. We would also like to thank Esra Alagoz, Aysegul Bakar, Sarah Chu, Yoonsin Oh, David Simkins, and, Bei Zhang for their assistance in data collection in Study Three and Study Four.

# Group Awareness of Social and Cognitive behavior in a CSCL Environment

C.Phielix & F.J. Prins, Utrecht University, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands,
C.Phielix@uu.nl, F.J.Prins@uu.nl
P.A. Kirschner, Open University, P.O. Box 2960, 6401 DL Heerlen, The Netherlands,
P.A.Kirschner@ou.nl

**Abstract:** This study investigated whether a peer feedback tool and a reflection tool would enhance group performance in a computer-supported collaborative learning environment. The underlying assumption was that group performance can be positively influenced by making group members aware of how their social and cognitive behavior is perceived by themselves, their peers, and the group as a whole. Participants were 120 fourth-year high school students working, with or without the tools, in dyads, triads and groups of 4 on a collaborative writing task. Results show that groups using tools perceived their team as being better developed, experienced higher levels of group satisfaction and lower levels of conflicts, than groups not using the tools. Results demonstrate that awareness, stimulated by peer assessment and reflection tools, enhances the social performance of a CSCL-group.

## Introduction

Collaborative learning, often supported by computer networks (computer supported collaborative learning, CSCL) is enjoying considerable interest at all levels of education. Collaborative learning, defined as the "mutual engagement of participants in a coordinated effort to solve the problem together" (Rochelle & Teasley, 1995, p. 70) has, among other things, been found to enhance the learners' cognitive performance (Johnson & Johnson, 1999) and to stimulate them to engage in knowledge construction (Stahl, 2004). The rapid development of information and communication technologies (ICT), has led to many computer applications (e.g., e-mail, chat, discussion forums, video conferencing, simulations, 3-D models, visualizations and external representations) which have proven to be useful tools or widgets to support collaborative learning (Janssen, Erkens, Kanselaar, & Jaspers, 2007).

Several researchers report cognitive and social benefits for groups in CSCL environments as compared to contiguous (i.e., face-to-face) groups. With respect to cognitive aspects of collaboration, researchers have found that students working in CSCL-environments report higher levels of learning (Hertz-Lazarowitz & Bar-Natan, 2002), make higher quality decisions, deliver more complete reports, participate more equally (Fjermestad, 2004; Janssen, Erkens, Kanselaar, & Jaspers, 2007), and engage in more complex, broader, and challenging discussions (Benbunan-Fich, Hiltz, & Turoff, 2003) than do students working face-to-face. With respect to social aspects, students working in CSCL-environments report higher levels of satisfaction compared to students in contiguous groups (Fjermestad, 2004).

There are, however, also contradictory results. Concerning the cognitive aspects of collaboration, students working in CSCL-environments sometimes perceive their discussions as more confusing (Thompson & Coovert, 2003), less productive (Straus, 1997; Straus & McGrath, 1994) and needing more time to reach consensus and make decisions (Fjermestad, 2004) than students working face-to-face. Also, students in CSCL-environments have been found to show lower levels of participation (Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003), and to experience higher levels of conflict (Hobman, Bordia, Irmer, & Chang, 2002), lower levels group cohesiveness (Straus, 1997; Straus & McGrath, 1994) and lower levels of satisfaction (Baltes, Dickson, Sherman, Bauer, & LaGanke, 2002). In other words, students working in CSCL-environments do not always reach their full potential. Two important reasons for the disparity between the potential of groups working in CSCL-environments and their performance lies in (1) the design of the CSCL-environment, and (2) the social and cognitive behavior of the group members.

With respect to design, CSCL environments often concentrate on functionality, focussing on the cognitive processes needed to accomplish a task and/or solve a problem (Kreijns & Kirschner, 2004). These functional CSCL environments coerce (Kirschner, Beers, Boshuizen, & Gijselaars, 2008) group members to limit their actions to cognitive processes to the detriment of socio-emotional processes. These socio-emotional processes, which are the basis for group forming and group dynamics, are essential for developing strong social relationships, strong group cohesiveness, feelings of trust, and a sense of community among group members (i.e., for creating a sound social space). Without such a sound social space, the group will not reach its full potential (Jehng, 1997). Groups in CSCL environments that lack social functionalities will ultimately perform poorly (e.g., Kreijns & Kirschner). For instance, despite technological advances, most CSCL still use text-based computer mediated communication (CMC) systems based on email, chat and/or discussion boards, which cannot easily convey visual nonverbal cues (Kreijns, Kirschner, & Jochems, 2003). The absence of these cues can cause specific problems for effective communication and interaction between group members since this removes possibilities for exchanging socio-emotional and affective information, and decreases information about group members' presence, self-image, attitudes, moods, actions and reactions (Short, Williams, & Christie, 1976). According to Short et al. the functions of these nonverbal cues are in some way related to forming, building or maintaining social relationships. Therefore, CMC can have negative effects on impression formation and group members' social behavior (e.g. Garton & Wellman, 1995; Walther, Anderson, & Park, 1994).

Second, group members form interpersonal perceptions during interaction (Kenny, 1994). Based on what they see and experience, they form impressions (e.g., norms, values, beliefs) about themselves, the group, other group members, and what the other group members think of them. These self-, other- and meta-perceptions are based on the cognitive behaviors (e.g., productivity) and social behaviors (e.g., dominance and friendliness) that they perceive during interaction. Based upon these perceptions, group members determine their own social and cognitive behavior, and develop social relationships with each other. However, research has shown that group members' perceptions of their own performance (i.e., self-perception) and of group performance are generally unrealistically positive, resulting in an illusion of group productivity (Stroebe, Diehl, & Abakoumkin, 1992). This tendency to believe that their group is performing effectively, while it often is not, can result in a reduction of effort by group members, a phenomenon also known as social loafing (Williams, Harkins, & Latané, 1981), which further undermines the groups' social and cognitive performance. However, group members are often not aware that they are loafing, or are unwilling to admit to it (Karau & Williams, 1993).

To overcome this obstacle to social and cognitive performance, CSCL environments can be augmented with computer tools or widgets that act as social contextual facilitators relevant for the learner's social interaction (Kirschner, Strijbos, Kreijns, & Beers, 2004). These tools, also known as 'social affordance devices', can positively affect social and cognitive performances in a CSCL environment (Kirschner, et al., 2004). Two operationalizations of such tools are used in this research, namely a peer feedback tool to make group members aware of the social and cognitive behavior of themselves, their peers, and how this is perceived by others, and a reflection tool to stimulate group members to reflect upon their individual behavior, why their peers see them the way they do, and to also reflect collaboratively (i.e., co-reflect) on the performance of the group as a whole. The aim of these tools is to make group members aware of their social and cognitive behavior and to enhance their social and cognitive performance and that of the group. The next sections deal with aspects central to these tools, namely peer feedback and reflection.

## Peer feedback

Feedback can be described as information provided to an individual to increase performance (Kluger & DeNisi, 1996). In a learning situation, this information can come from many sources such as teachers, computers, fellow students, and so forth. This study centers on peer feedback; information provided by fellow learners which is intended to increase performance. This information can be provided on the outcome performance (i.e., outcome feedback), or on how one is performing (i.e., process feedback). Feedback can be given by individuals or groups, and can also be received by individuals or groups. In this study, peers provide process feedback at individual and group level, in order to enhance interpersonal behavior. It is expected that enhancement of interpersonal behavior will have a positive effect on a group's social performance (Geister, Konradt, & Hertel, 2006; McLeod & Liker, 1992; Phielix, Prins & Kirschner, in press), as well as an indirect positive effect on a group's cognitive performance (Kreijns, Kirschner, & Jochems, 2003).

Process feedback can include cognitive or task-related information (e.g., task behaviors, actions and strategies), or social or non-task related information, such as information about interpersonal behavior (e.g., dominance and friendliness) or teamwork (Geister, Konradt, & Hertel, 2006). McLeod and Liker (1992) found that process feedback at the group level on the interpersonal behavior of student group members, such as dominance and group orientedness, changed the dominance behavior of individual group members. Two other studies investigating individualized peer feedback on interpersonal behavior of group members (e.g., communication and collaboration), found that such feedback led to increased cooperation, communication, satisfaction and motivation in group members (Dominick, Reilly, & McGourty, 1997; Druskat & Wolff, 1999).

Based upon these ideas, an individualized peer feedback tool (Radar) was developed and studied in which group members individually provide information about the social and cognitive behavior of themselves, their peers, and the group as a whole. The premise behind this tool is that it will positively alter the social and cognitive behavior of individuals and group. Because group members tend to overestimate their social and cognitive behavior (Kenny, 1994), this peer feedback tool also included the peer perspectives on the social and cognitive behavior of themselves, their peers and the group as a whole. This information should be gathered by use of a self and peer assessment, and based on specific traits because there is strong evidence that peer perceptions are formed by unconscious or tacit 'rating' of other group members on several traits, such as 'dominance', 'friendliness' and 'reliability' (Brok, Brekelmans, & Wubbels, 2006).

## Reflection

Simply providing group members' with information on their cognitive and social behavior is not enough to positively alter their behavior (Prins, Sluijsmans, & Kirschner, 2006). Group members also need to process this information and ask themselves whether they understand, accept, and agree with the feedback. In other words, they must reflect upon the feedback. Reflection is the intellectual and affective activities individuals engage in to explore their experiences (e.g., behaviors, ideas, feelings) in order to reach new understandings and appreciations (Boud, Keogh, & Walker, 1985). The feedback receiver needs to be challenged to reflect on his/her own performance, and determine whether the feedback provides clues for behavioral change (Prins, Sluijsmans, & Kirschner, 2006). Therefore, it is expected that peer feedback in combination with reflection will even be more effective than feedback alone (e.g., Schön, 1987).

According to Boud, Keogh, and Walker (1985), reflection can lead to new perspectives on experience, changes in behavior, readiness for application, and commitment to action. Therefore, reflection on peer feedback should make group members more aware of their own individual behavior, how their behavior affects others, and whether they should alter their behavior. Awareness can be defined as the "understanding of the activities of others, which provides a context for your own activity" (Dourish & Bellotti, 1992, p. 107). In order to make the group aware of its behavior, group members need to reflect collaboratively (co-reflect) on their cognitive and social performance. Co-reflection is defined as "a collaborative critical thinking process involving cognitive and affective interactions between two or more individuals who explore their experiences in order to reach new intersubjective understandings and appreciations" (Yukawa, 2006; p. 206).

Based upon these ideas, a shared reflection tool (Reflector) was developed and studied in which group members individually reflect and provide information on (1) their own individual perspective on their personal performance, (2) differences between their self perception and the perception of their peers concerning their personal performance, (3) whether they do or do not agree with the perceptions of their peers concerning their personal performance, and (4) their own individual perspective on group performance. Because group performance is determined by the individual effort of all group members, this tool also (5) stimulates group members to collaboratively reflect (co-reflect) on the group performance and reach a shared conclusion about this. Based on their shared conclusion, group members will (6) set goals in order to improve group performance.

## Research Questions

This study investigated whether a peer feedback tool and a reflection tool would enhance group performance in a CSCL-environment. To this end, an existing CSCL-environment was augmented with two independent, but complementary, tools. The first was an individualized peer feedback tool - Radar - which was meant to stimulate and provide group members with information about the social and cognitive behavior of themselves, their peers, and the group as a whole. This information was presented from both the perspectives of the group members themselves (i.e., self perceptions), their peers (i.e., peer perceptions) and the group as a whole. The second tool was a shared reflection tool - Reflector - which was meant to stimulate group members to reflect on and provide information about their own behavior and how this behavior was perceived by their peers, their personal perspectives on the group's performance, as well as to co-reflect on the group performance and reach shared understanding on this.

The following research questions will be addressed:

1) Do groups with Radar and Reflector perform better socially than groups without Radar and/or Reflector? In other words, do groups using Radar and Reflector develop better, have higher group satisfaction, experience lower levels of group conflict, and a have more positive attitude towards collaborative problem solving than groups without Radar and/or Reflector?
   Expected is that both Radar and Reflector will positively affect the social behavior in the group, leading increased social performance of the group. A combination of the both tools should be most effective.
2) Do groups with Radar and Reflector perform better cognitively than groups without Radar and/or Reflector? In other words, do groups with Radar and Reflector produce a group product of higher quality than groups without Radar and/or Reflector?
   Expected is that both Radar and Reflector will positively affect the social behavior in the group and that this should indirectly lead increased cognitive performance of the group, A combination of both tools should be most effective.

## Method

### Participants

Participants were 120 fourth-year students (66 male, 54 female) from an academic high school in The Netherlands. Students came from four classes and were enrolled in the second stage of the pre-university education track which encompasses the final three years of high school. Prior to the experiment, the participating students were randomly assigned by the teacher to dyads, triads and groups of four, and randomly assigned by the teacher to one of the three conditions (see Design). Therefore, group compositions were heterogeneous in ability and gender.

## Design

For this study two experimental conditions and one control condition were used. The first experimental condition ($n$ = 69) received the tools at the beginning (T1), halfway (T2) and at the end (T3) of the collaboration process. The second experimental condition ($n$ = 24) received the tools halfway (T2) and at the end (T3). The control condition ($n$ = 27) did not received tools during collaboration but only completed them at the end (T3).

## Measures

*Cognitive performance.* The grade given to the groups' collaborative writing task (i.e., the essay) was used as a measure of cognitive performance. The essays were graded by two researchers, both experienced in grading essays. The inter-rater reliability was high (Cronbach's $\alpha$ = .86).

*Social performance.* To measure social performance, previously validated instruments (Strijbos, Martens, Jochems, & Broers, 2007) were translated into Dutch and transformed into 5-point Likert scales (1 = totally disagree, 5 = totally agree). The Team Development scale ($k$ = 13, $\alpha$ = .88) provides information on the perceived level of group cohesion. The Group-process Satisfaction scale ($k$ = 9, $\alpha$ = .79) provides information on the perceived satisfaction with general group functioning. The Intra-group Conflicts scale ($k$ = 10, $\alpha$ = .88) provides information on the perceived level of conflict between group members. The Attitude towards Collaborative Problem Solving scale ($k$ = 9, $\alpha$ = .78) provides information on the perceived level of group effectiveness and how group members felt about working and solving problems in a group.

## Task and procedure

The students collaborated in dyads and groups of three or four on a collaborative writing task in sociology. Every student worked at a computer. Each group had to write an essay on a highly relevant current-events topic. Prior to this collaborative writing task, students collaborated for one month choosing the topic, searching for relevant sources, writing a short paper and giving a class presentation. Therefore, the sources required to write the essay were available for all groups. The collaborative writing task consisted of three 45-minute sessions over a period of one week. The groups collaborated in a CSCL environment called Virtual Collaborative Research Institute (VCRI; Jaspers, Broeken, & Erkens, 2002), which is a groupware program designed to support collaborative learning on research projects and inquiry tasks. VCRI will be further described in the Instruments section. Students were instructed to use VCRI to communicate with the other group members and to make complete use of the tools for peer feedback and reflection when the experimental condition allowed this. Students received content information and definitions regarding the six traits on which they had to assess themselves and their peers. Students were told that they had three lessons to complete the task, that it would be graded by their teacher, and that it would affect their grade for the course. The introduction to the task stressed the importance of working together as a group and pointed out that each individual group member was responsible for the successful completion of the group task. To successfully complete the task, all group members had to participate.

While groups used the tools, groups without tools continued working on their collaborative writing task. Time-on-task (writing the essay) was equal for all conditions. At the end of the final session (T3), the peer feedback tool and reflection tool became available for all conditions so that all participants could assess their peers and reflect on their behaviors. Finally, all participants completed a 30-item questionnaire measuring the social performance of the group.

## Instruments

### Virtual Collaborative Research Institute (VCRI)

The Virtual Collaborative Research Institute (VCRI) is a groupware program that supports collaborative working and learning on research projects and inquiry tasks (Jaspers, Broeken, & Erkens, 2004). VCRI contains more than 10 different tools, but only 5 were used for this experiment (see Figure 1). The Co-Writer (top left) is a shared word-processor for writing a group text. Using the Co-Writer, students can simultaneously work on different parts of their texts. The Chat tool (top center) is used for synchronous communication between group members. The chat history is automatically stored and can be re-read by participants at any time. Notes (bottom right) is a note pad which allows the user to make notes and to copy and paste selected information. Radar for peer feedback (bottom left) and Reflector for reflection (top right) will be described in the following sections. Windows of the available tools are automatically arranged on the screen, when students log on to the VCRI.

### Peer assessment tool (Radar)

VCRI was augmented with a peer feedback tool for stimulating and facilitating information of group members' social and cognitive behavior. This information is visualized in a radar diagram; therefore the peer feedback tool is named 'Radar' (see Figure 2). Radar provides users with anonymous information on how their cognitive and social behavior is perceived by themselves, their peers, and the group as a whole. The information gathered is based on specific traits that have been found to tacitly affect how one 'rates' other people (Brok, Brekelmans, & Wubbels, 2006). Radar provides information on six traits that are important for assessing behavior in groups. Four are related to social or interpersonal behavior, namely (1) influence; (2) friendliness; (3) cooperation; (4) reliability; and two are related to cognitive behavior, namely (5) productivity and (6) quality of contribution. These traits are derived from studies on interpersonal perceptions, interaction, group functioning, and group effectiveness (e.g., Bales, 1988; Brok, Brekelmans, & Wubbels; Kenny, 1994). These traits, as well as the reasons for their choice, are discussed in Phielix, Prins, and Kirschner (in press).

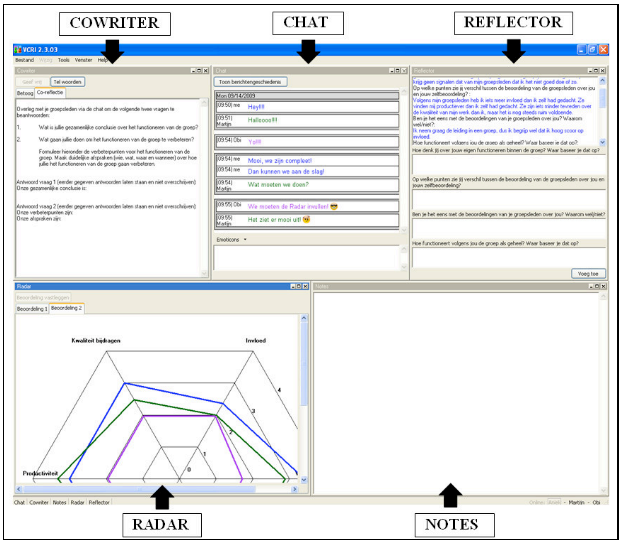


Figure 1. Screenshot of VCRI

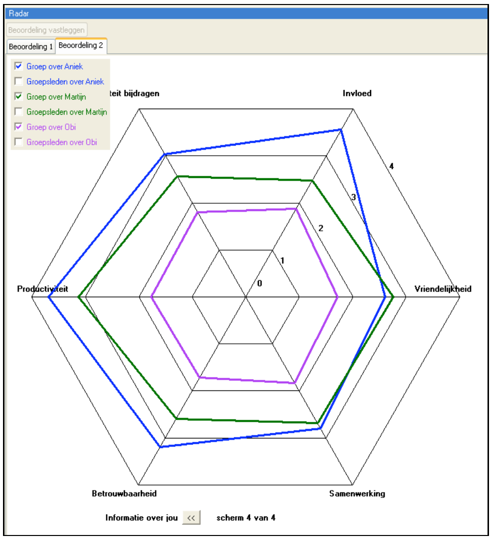


Figure 2. Output group assessment

In Radar, group members are both assessors and assessees. As assessor, to-be-assessed peers in the group can be selected and her/his profile will appear as dotted lines in the center circle of the radar diagram. Each group member is represented by a specific color. The assessor rates her/himself and all of the other group members on each of the six traits using a continuous scale ranging from 0 to 4 (0 = none - 4 = very high). Every range, (e.g., from 0 to 1) contains 10 points, so every scale contained 40 points of assessment. To make sure that all assessors interpret the six traits in the same way, assessors saw a text balloon with content information and definitions when they moved the cursor across one of the traits in the tool. For example, when the assessor moves the cursor across 'influence' a balloon pops up with the text 'A high score on influence means that this person has a big influence on what happens in the group, other group members behavior, and the form and content of the group product (the essay)'.

For groups of 3 and 4 members, the assessment is anonymous. Group members can see the assessments of the other group members, but not who entered the data. To stimulate students to complete Radar, they can only gain access to the individual and average assessments of their peers after they have completed the assessment themselves. When all group members have completed their self assessments and peer assessments, two modified radar diagrams become available. The first - Information about yourself - shows the output of the self assessment (e.g., Chris about Chris) along with the average scores of the peer assessments of her/him (e.g., Group about Chris). The self assessment is not taken into account for computing the average scores. To provide more information about the variance in the average score of their peer assessment, students can also choose to view the individual peer assessments about their own behavior (e.g., Group members about Chris). The second - Information about the group (see Figure 2) - represents the average scores of the group members, so that group members can get a general impression about the functioning of the group.

All group members are represented as a solid line in the diagram, each with a different color. The student can include or exclude group members from the diagram by clicking a name in the legend.

## Reflection tool (Reflector)

VCRI was also augmented with a reflection tool (Reflector) containing five reflective questions designed to stimulate reflection on different aspects of the group processes taking place. The questions were:

1. What is your opinion on how you functioned in the group? Give arguments to support this.
2. What differences do you see between the assessment received from your peers and your self assessment?

3. Why do you or do you not agree with your peers concerning your assessment?
4. What is your opinion on how the group is functioning? Give arguments to support this.
5. What does the group think about its functioning in general? Discuss and formulate a conclusion that is shared by all the group members.
6. Set specific goals (who, what, and when) in order to improve group performance.

The first four questions are completed in the Reflector, with completion indicated by clicking an 'Add'-button. This allows the student to share her/his answers with the rest of the group and allows her/him to see the answers of the others. Students can only gain access to the answers of their peers after they have added their own answers so as not to be influenced by one another. The fifth and sixth questions are completed in Co-Writer, in a specific section named Co-Reflection, which allows writing a 'shared' conclusion and formulating goals. The responses made by the students in the Reflector are not scored or evaluated.

## Data Analyses

To examine whether Radar and Reflector lead to higher social performance, a one way between-groups ANOVA (two-tailed) with planned comparisons is conducted with the dependent variables 'team development', 'group satisfaction', 'level of group conflicts', and 'attitude towards collaborative problem solving', as measured by the questionnaire at the end of the experiment.
To examine whether Radar and Reflector lead to higher cognitive performance, a one way between-groups ANOVA (two-tailed) with planned comparisons is conducted with the grade on the essay as dependent variable.

## Results

*Impact of tools on social performance.* A one way between-groups ANOVA (two-tailed) with planned comparisons was conducted to compare scores on 'team development', 'group satisfaction', 'level of group conflicts', and 'attitude towards collaborative problem solving', across each of the three conditions. Groups in Condition 1 used the tools from the beginning (T1) of collaboration process until the end (T3); Groups in Condition 2 received and used the tools halfway (T2) and at the end (T3); and Groups in Condition 3 did not use the tools during collaboration, but only completed them at the end (T3). Except where noted, tests were two-sided. The rule of thumb (Cohen, 1988) for effects sizes ($\eta^2$) was small $\geq .01$, medium $\geq .06$, and large $\geq .14$. Table 1 shows means and standard deviations for social performance scales per condition.

Table 1. Means and Standard Deviations for Social Performance Scales per Condition

| | Condition | *N* | *M* | *SD* |
|---|---|---|---|---|
| Team development | 1 – tools available at T1, T2 and T3 | 63 | 4.06 | .57 |
| | 2 – tools available at T2 and T3 | 24 | 3.47 | .61 |
| | 3 – tools available at T3 | 25 | 3.80 | .37 |
| Group satisfaction | 1 – tools available at T1, T2 and T3 | 62 | 3.98 | .56 |
| | 2 – tools available at T2 and T3 | 24 | 3.56 | .64 |
| | 3 – tools available at T3 | 25 | 3.71 | .59 |
| Intra-group conflict | 1 – tools available at T1, T2 and T3 | 62 | 1.95 | .59 |
| | 2 – tools available at T2 and T3 | 24 | 2.41 | .66 |
| | 3 – tools available at T3 | 25 | 2.23 | .50 |
| Attitude towards collaborative problem solving | 1 – tools available at T1, T2 and T3 | 62 | 3.79 | .59 |
| | 2 – tools available at T2 and T3 | 24 | 3.56 | .61 |
| | 3 – tools available at T3 | 25 | 3.65 | .55 |

As expected, groups that used the tools during the complete collaboration process (Condition 1), perceived their team as being better developed, $F(1, 68) = 6.10$, $p = .02$, partial $\eta^2 = .16$, experienced higher levels of group satisfaction, $F(1, 108) = 3.83$, $p = .05$, partial $\eta^2 = .09$, and experienced lower levels of conflicts, $F(1, 108) = 4.07$, $p = .05$, partial $\eta^2 = .10$, than students not using the tools (Condition 3).
Compared to groups that received the tools halfway (Condition 2), groups in Condition 1 perceived their team as being better developed, $F(1, 68) = 17.02$, $p = .00$, partial $\eta^2 = .16$, experienced higher levels of group satisfaction, $F(1, 108) = 9.02$, $p = .00$, partial $\eta^2 = .09$, experienced lower levels of conflicts, $F(1, 108) = 10.68$, $p = .00$, partial $\eta^2 = .10$, and had a more positive attitude towards collaborative problem solving, $F(1, 108) = 2.77$, $p = .05$, partial $\eta^2 = .03$ (one-tailed).
Compared to groups in Condition 3, groups in Condition 2 perceived their team as being less developed, $F(1, 38) = 5.40$, $p = .03$, partial $\eta^2 = .16$, compared to students not using the tools (condition 3).

Table 2. Means and Standard Deviations for Cognitive Performance per Condition

| | | Cognitive performance (grade essay) | | | |
|---|---|---|---|---|---|
| Condition | *N* | *M* | *SD* | *Min* | *Max* |
| 1 – tools available at T1, T2 and T3 | 69 | 6.81 | .71 | 4.0 | 8.5 |
| 2 – tools available at T2 and T3 | 21 | 6.54 | 1.04 | 4.5 | 8.5 |
| 3 – tools available at T3 | 27 | 6.36 | 1.61 | 4.0 | 8.5 |

*Impact of tools on cognitive performance.* Table 2 shows means and standard deviations for performance per condition as measured by the essay grades. No significant effects of Radar and Reflector were found.

## Discussion and Conclusion

The first aim of this study was to examine whether the use of a peer feedback tool (Radar) and reflection tool (Reflector) would lead to higher social performance, measured by: team development, group satisfaction, level of group conflict, attitude towards collaborative problem solving. As expected, groups using the tools perceived their team as being better developed, experienced higher levels of group satisfaction and lower levels of conflicts, than groups not using the tools or groups using the tools since halfway the collaboration process. Groups using the tools also experienced a more positive attitude towards collaborative problem solving, than groups receiving the tools halfway. Surprisingly, groups receiving the tools halfway perceived their team as being less developed compared to groups not using the tools. An explanation could be that Radar and Reflector made group members aware of their unrealistic positive perception on team development.

The second aim was to examine whether the use of Radar and Reflector would lead to higher cognitive performance, measured by the grade given to the essays. No significant effects of Radar and Reflector were found for grade given to the essays. The lack of a significant effect is probably due to the short period of time in which the groups had to collaborate in order to accomplish the task. Therefore, further studies will examine the effects of Radar and Reflector during a longer period of time (i.e., three months).

In conclusion, the effects of Radar and Reflector are very promising. They show that social group performance in CSCL environments, such as team development, group satisfaction, and level of group conflicts, can be enhanced by adding this easy to complete and easy to interpret peer feedback tool and reflection tool.

## References


Bales, R. F. (1988). A new overview of the SYMLOG system: Measuring and changing behavior in groups In R. B. Polly, A. P. Hare, & P. J. Stone (Eds.), *The SYMLOG practitioner: Applications of small group research* (pp. 319-344).

Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes, 87*(1), 156-179.

Benbunan-Fich, R., Hiltz, S. R., & Turoff, M. (2003). A comparative content analysis of face-to-face vs. asynchronous group decision making. *Decision Support Systems, 34*, 457-469.

Brok, P. den, Brekelmans, M. & Wubbels, Th. (2006). Multilevel issues in studies using students' perceptions of learning environments: the case of the Questionnaire on Teacher Interaction. *Learning Environments Research, 9*, 199-213.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The effects of peer feedback on team member behavior. *Group & Organization Management, 22*, 508-525.

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. In M. Mantel, & R. Baecker (Eds.), *Proceedings of the 1992 ACM conference on Computer-supported cooperative work* (pp. 107–114). New York: ACM Press.

Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology, 84,* 58-74.

Fjermestad, J. (2004). An analysis of communication mode in group support systems research. *Decision Support Systems, 37*, 239-263.

Garton, L., & Wellman, B. (1995). Social impacts of electronic mail in organizations: a review of the research literature. In B. R. Burleseon (Ed.), *Communication yearbook (Vol. 18)* (pp. 438–453). Thousand Oaks, CA: Sage.

Geister, S., Konradt, U., & Hertel, G. (2006). Effects of Process Feedback on Motivation, Satisfaction, and Performance in Virtual Teams. *Small Group Research*, *37*, 459-489.

Hertz-Lazarowitz, R., & Bar-Natan, I. (2002). Writing development of Arab and Jewish students using cooperative learning (CL) and computer-mediated communication (CMC). *Computers & Education, 39*, 19-36.

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning. *Computers & Education, 49*, 1037-1065.

Jaspers, J., Broeken, M., & Erkens, G. (2004). *Virtual Collaborative Research Institute (VCRI) (Version 2.0).* Utrecht: Onderwijskunde Utrecht, ICO/ISOR.

Jehng, J.J. (1997). The psycho-social processes and cognitive effects of peer-baded collaborative interactions with computers. *Journal of Educational Computing Research, 17(1), 19-46*

Johnson, D. W., & Johnson, R. T. (1999). *Learning together and alone: Cooperative, competitive, and individualistic learning* (5th ed.). Boston: Allyn and Bacon.

Karau, S., & Williams, K. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*, 681-706.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.

Kirschner, P. A., Beers. P. J., Boshuizen, H. P. A., & Gijselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior, 24*, 403-420.

Kirschner, P., Strijbos, J., Kreijns, K., & Beers, P. J. (2004). Designing electronic collaborative learning environments. *Educational Technology Research and Development*, *52*(3), 47–66.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.

Kreijns, K. & Kirschner, P. A., (2004). Determining sociability, social space and social presence in (a)synchronous collaborating teams. *Cyberpsychology and Behavior, 7, 155-172.*

Kreijns, K., Kirschner, P. A. & Jochems, W. (2003a). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior, 19, 335-353.*

Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2003). Patterns of participation and discourse in elementary students' computer-supported collaborative learning. *Learning and Instruction, 13*, 487-509.

McLeod, P. L., & Liker, J. K. (1992). Process feedback in task groups: An application of goal setting. *Journal of Applied Behavioral Science, 28*, 15-52.

Phielix, C., Prins, F. J., Kirschner, P. A. (in press). Awareness of group performance in a CSCL environment: Effects of peer feedback and reflection. *Computers in Human Behavior.*

Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Rochelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). New York: Springer-Verlag.

Schön, D. A. (1987). *Educating the reflective practitioner.* San Francisco, CA: Jossey-Bass.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* London: John Wiley & Sons.

Stahl, G. (2004). Groupware goes to school: Adapting BSCW to the classroom. *International Journal of Computer Applications in Technology, 19*(3/4), 1–13.

Straus, S. G. (1997). Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups. *Human-Computer Interaction, 12*, 227-266.

Straus, S. G., & McGrath, J. E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. *Journal of Applied Psychology, 79*(1), 87-97.

Strijbos, J.W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2007). The effect of functional roles on perceived group efficiency during computer-supported collaborative learning: a matter of triangulation. *Computers in Human Behavior, 23*, 353–380.

Stroebe, W., Diehl, M. & Abakoumkin, G. (1992). The illusion of group effectivity. *Personality and Social Psychology Bulletin, 18,* 643-650.

Thompson, L. F., & Coovert, M. D. (2003). Teamwork online: The effects of computer conferencing on perceived confusion, satisfaction and postdiscussion accuracy. *Group Dynamics, 7*, 135-151.

Walther, J. B., Anderson, J. F., & Park, D. (1994). Interpersonal effects in computer-mediated interaction: a meta-analysis of social and anti-social communication. *Communication Research, 19*, 460–487.

Williams, K. D., Harkins, S. G., & Latané, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology, 40*, 303-311.

Yukawa, J. (2006). Co-reflection in online learning: Collaborative critical thinking as narrative. *Journal of Computer-Supported Collaborative Learning, 1*, 203-228.

# The Impact of a Media-Rich Science Curriculum on Low-Income Preschoolers' Science Talk at Home

William R. Penuel, SRI International, 333 Ravenswood, Menlo Park CA 94025, william.penuel@sri.com
Lauren Bates, Education Development Center, 96 Morton, NY, NY 10014, lbates@edc.org
Shelley Pasnik, Education Development Center, 96 Morton, NY, NY 10014, sp@edc.org
Eve Townsend, Education Development Center, 96 Morton, NY, NY 10014, etownsend@edc.org
Lawrence P. Gallagher, SRI International, 333 Ravenswood, Menlo Park CA 94025, lawrence.gallagher@sri.com
Carlin Llorente, SRI International, 333 Ravenswood, Menlo Park CA 94025, carlin.llorente@sri.com
Naomi Hupert, Education Development Center, 96 Morton, NY, NY 10014, nhupert@edc.org

**Abstract:** While research suggests that educational television shows can contribute positively to a range of developmental outcomes for preschoolers, few preschool curricula make use of digital content to teach science. This study explored the impact of a curriculum that integrated hands-on activities with digital content from two public television shows aimed at introducing preschoolers to science, *Peep and the Big Wide World* and *Sid the Science Kid.* Impact was measured using parent reports of 398 low-income children's science talk using a random assignment design, where preschool teachers were assigned to implement either the media-rich science curriculum or a literacy curriculum. Results indicated that the science curriculum had a positive impact on caregivers' reports of children's talk about science. Though the study's outcome measure is an indirect measure of learning, the study suggests the potential for this and other media-rich curricula for introducing preschoolers to science.

## Context for the Study

Learning scientists are increasingly interested in preschool children's science learning. Learning scientists have examined young children's conceptual development (Gelman & Lucariello, 2002; Gelman & Brennamen, 2004), the roles that caregivers play in encouraging young children's talk about science at home that encourages curiosity and fosters the development of explanations (Crowley et al., 2001), and the effects of specific curricula that target preschoolers in developing children's inquiry skills and conceptual understanding (Clark-Chiarelli, Gropen, Chalufour, Hoisington, & Eggers, 2009). To date, however, there have been few studies exploring the role that digital media can play in supporting young children's science learning or that focus specifically on science learning of low-income children. This gap is noteworthy, since a significant focus of learning sciences research has been on the role technology can play in supporting science learning in K-12 settings (Pea & Collins, 2008) and on the requirements for enacting inquiry approaches to teaching science in low-income communities (Songer, Lee, & Kam, 2002).

There are several obstacles to designing and undertaking such studies in preschool settings. First, the use of technology in preschools has been controversial. Critics have argued that adults should not promote the use of video or computers by young children, arguing that these technologies make children passive and reduce their opportunities to learn from interacting with adults and peers (Healy, 2003). In preschools where staff members believe technology is likely to harm children, it may be difficult to recruit study participants. In addition, curricula that use technology can also be hard for teachers to implement, since many preschools have limited access to technology and teacher training opportunities (Davidson, Fields, & Yang, 2009). Thus, fielding a study of media-rich curriculum materials would require researchers to provide extensive technical and professional support. Third, there are few standards-based outcomes for researchers to use. Those few that have some evidence of validity and reliability are only available to those researchers involved in their development. As a result, researchers wishing to study outcomes will have to develop measures that are instructionally sensitive, yet not so closely aligned to curriculum content as to have limited value in persuading others of the objectivity of the measure.

This study investigates whether low-income preschool-aged children can learn from a media-rich curriculum that combines hands-on investigations with digital content from public television programs and games. Researchers at Education Development Center Inc. designed the curriculum used in the study, which was part of the *Ready to Learn Initiative*, a program to develop educational television programming and outreach activities that increase school readiness for 2- to 8-year-old children living in low-income households. In 2005, the U.S. Department of Education awarded one of two Ready to Learn programming grants to the Corporation for Public Broadcasting (CPB) and PBS. The award required an impact evaluation of the *Ready to Learn* Initiative's target audience of low-income children.

## Theoretical Background

The premise behind the curriculum in this study is that preschool teachers can implement a curriculum that integrates digital content from public television preschool science programs with hands-on activities in ways that promote early science learning. The curricular design is grounded in the literature on early science learning and on how and when young children can learn from digital media.

### Approaches to Promoting Early Science Learning

As is true for learning across a variety of domains and ages, young children's prior experiences are important for learning. Because children have every day experiences with life science and earth science phenomena, many preschool science curricula draw from these fields. For instance, Rule and Guggenheim's (2007) science curriculum centers on clay partly because young children often have prior experience working with clay. The curriculum progresses from using art clay to experimenting with mud, clay, and sands. Gelman and Brenneman's (2004) curriculum embeds the scientific method and vocabulary into lessons devised with familiar objects like the apples, seeds, baby pictures, and shoes of various sizes, all of which demonstrate the concepts of growth and change. The idea behind both curricula is young children can be introduced to science concepts by using familiar objects and phenomena.

Hands-on investigations of these phenomena enable students to develop critical skills for inquiry, including making observations and predictions. Because hands-on activities make observations concrete and more comprehensible to preschoolers, science curricula routinely call for children to experiment with objects by using the senses to explore and manipulate them. Many preschool science curricula teach students to record observations with drawings and then make predictions based on the observations (e.g., Chalufour & Worth, 2003). Long-term activities also promote deeper exploration of a topic by allowing preschoolers to observe, predict, verify predictions, and reflect on observed changes over time (Gelman & Brenneman, 2004). Such investigations also support the earliest stage of scientific thinking, false belief understanding, which occurs when a child realizes her observations or beliefs may be different from others' and that beliefs may even be false (Wellman, Cross, & Watson, 2001). False belief understanding is the first step towards the ability to falsify a theory (Kuhn & Pearsall, 2000).

In early childhood, encounters with science can help children cultivate students' interest in science and develop ways for talking about science and learn scientific vocabulary, two skills that correspond to important strands of proficiency in science (National Research Council, 2009). According to French (2004), children's interest in science activities can support the formation of mental representations of complex phenomena and promote communication about these phenomena. For example, preschoolers can verbally describe the inside and outside of an apple, describing the apple's color, texture, temperature, and lack of sound while the teacher records the information on a chart (Gelman & Brenneman, 2004).

To date, little research has focused on science learning for low-income preschoolers; an exception is a recent pair of studies focused on children's science learning in Head Start by Greenfield and colleagues (Greenfield et al., 2009). In one study, the researchers compared the developmental scores of Head Start children using the Galileo observational system (which rates children's skills in a range of domains). They found that in a cohort of more than 2,000 children, children's ratings and gains in science were significantly lower than in the other 7 domains measured (e.g., literacy). In another study, they found that many teachers reported low perceived self-efficacy with respect to teaching science and had trouble integrating science into their schedule. A third study was a quasi-experimental study of the ECHOS curriculum, which uses a combination of direct instruction and guided discovery to teach science skills. Relative to a control group, students in classrooms implementing the ECHOS curriculum were rated higher on four of the eight domains of the Galileo observational system.

### Young Children's Learning Using Digital Media

Research on young children's learning from digital media has been an area where there is a long history of research, including for low-income students (for a review, see Pasnik, Strother, Shindel, Penuel, & Llorente, 2007). For example, in the earliest days of public television programming for children, producers used research to study children's engagement and learning (Fisch, Truglio, & Cole, 1999; Morrow, 2005). A pair of experimental studies conducted by the Educational Testing Service in the first and second years of production for *Sesame Street* (Ball & Bogatz, 1970; Bogatz & Ball, 1971) found that children encouraged to watch the show grew in skills targeted by the show and were rated as more ready for school by their teachers than students assigned to a control group. A limitation of these and many of the other studies of the effects of educational television, including those that employed random assignment designs, is that they have been conducted in laboratory settings (Thakkar, Garrison, & Christakis, 2006). The lack of field-based studies is significant because achieving similar results in field settings (e.g., homes, day care facilities, or early childhood education centers) requires that a coherent sequence of curricular activities be enacted, something care givers in these settings may find challenging.

With the increasing availability of computers came developments in computer-assisted instruction that targeted young children's learning, primarily in the area of literacy. For this age group that is not yet reading, software programs facilitate the reading process by having the computer speak letter sounds, phonemes, and words to students as they read or interact with the computer. An intervention designed to increase phonological awareness (Olson, Wise, Ring, & Johnson, 1997) is fairly typical of such programs. In their software program, one set of tasks asks children to change the onset or rime of a word presented on a screen to match what the computer says (e.g., if the computer shows "buzz" and the computer says "fuzz," students must replace the "b" with an "f"). Another set of tasks requires children to try spelling words. The computer pronounces students' spellings back to them, providing them with feedback they can use to adjust their spelling. Studies of the effectiveness of these programs (e.g., Foster, et al. 1994; Baker & Torgesen, 1995) have found strong, positive effects on children's phonological awareness.

More recently, a number of interventions have been developed based on the idea that different media can be used *synergistically* (Neuman, 1995) to enhance children's learning. Interventions designed to promote learning from media synergy encourage co-viewing (watching and interacting with the television program with a peer, parent, or teacher) coordinate media viewing with non-media activities such as listening to a story, reading a book with a parent, teacher, or older sibling, or working on a practice activity related to skills targeted in the intervention. An intervention described by Prince and colleagues (Prince et al, 2002) provides a representative example of a media synergy intervention in practice. Teachers in that study were provided with a comprehensive *Between the Lions* curriculum that included whole episodes of the program, books related to themes covered during the program and enrichment activities. Teachers participated in intensive, daylong workshops to familiarize themselves with the resources and learn strategies for using them to supplement the literacy curricula already in place in their schools. During the school year, participating preschool, kindergarten, and first grade children viewed at least two *Between the Lions* episodes, read a book related in some way to the content of the episode viewed, and then participated in a hands-on activity that reinforced the skill or theme stressed in the episode. This and other studies (e.g., Chambers, et al., 2006) have found small, but positive, effects on children's reading skills. Similarly, recent evaluations of preschool math curricula that include computer-based math games revealed significant student math achievement gains (Clements & Sarama, 2008; Starkey, Klein, & Wakeley, 2004).

## The Public Media-Produced Science Curriculum

The 10-week media-rich early science curriculum studied integrated video from educational television programs and associated online games with classroom activities to foster skills for later success in science learning. Researchers randomly assigned teachers to the early science or an early literacy condition. Teachers played a leading role in every aspect of the curriculum, guiding students through whole-group hands-on investigations, small-group experiences, and individual exploration, as well as mediating children's experiences with digital content through active co-viewing. The curriculum combined full episodes of *Sid the Science Kid* (produced by KCET/Los Angeles with Jim Henson Productions) and self-contained "focused viewing" segments from *Peep and the Big Wide World* (produced by WGBH Educational Foundation). Over the 10 weeks, the children participated in 25 hours of the science curriculum, including time spent actively watching TV episodes and segments, playing online games, and participating in small and large group activities.

### Integration of Different Media and Formats for Learning

Video from *Sid the Science Kid* and *Peep and the Big Wide World* anchored the curriculum activities. Teachers showed entire episodes of *Sid the Science Kid* to introduce a weekly in-class investigation, periodically interrupting the video to highlight key science vocabulary and ideas. Students also viewed self-contained segments from *Peep and the Big Wide World* that engaged target content, vocabulary, and skills. Online games created by the producers provided additional support for the curriculum by targeting the same instructional strands. Both programs are intended to develop the early scientific knowledge of preschool children, the target age for the study. Researchers also adapted the hands-on activities suggested on the programs' websites for inclusion in the curriculum. Teachers led hands-on activities in whole group, small group, and individual settings.

### Focal Skills

The curriculum addressed the development of four instructional strands in early science: science content, vocabulary, skills, and scientific thinking. In conjunction with students' prior knowledge, hands-on exploration, and observations, each strand is essential for later scientific development. These strands align with three of the National Research Council's recommended strands for science learning in informal environments: 1) Experience excitement, interest, and motivation to learn about phenomena in the natural and physical world; 3) Manipulate, test, explore, predict, question, observe, and make sense of the natural and physical world; and 5) Participate in scientific activities and learning practices with others, using scientific language and tools (National Research Council, 2009). Teachers guided children in exploring *science content* that was conceptually linked to "Transformation and Change," the topic of the 5 *Sid* episodes selected for the study. Instructional content addressed freezing, melting, growth, decay, reversible change, and irreversible change. Activities were based on everyday experiences easily observable with the five senses. Teachers taught preschoolers *science vocabulary* words related to these science concepts. They also learned vocabulary for science skills, such as observing, collecting, sorting, and investigating. Teachers supported children in learning the *science skills* needed to make observations using their senses, and record them in drawings. Long-term investigations allowed children to make predictions and check them over time. Children practiced *scientific thinking* by comparing their observations with those of their classmates, as well as by participated in memory games.

### Teacher-Led Viewing

The curriculum called for early childhood educators to engage children in active viewing of segments and episodes. When showing video to the children, teachers introduced the key content and vocabulary, paused the video to encourage active processing of information, and reflected on areas of learning embedded in the video. After viewing each week's *Sid* episode, teachers supported children in repeating the hands-on investigations performed by the characters, such as planting a seed to watch it grow.

### Opportunities for Repeated Practice

The curriculum sequence gave children multiple opportunities to develop and practice focal skills. They watched whole episodes and focused viewing segments several times and participated in repetitive teacher-led, small-group, and individual activities. Review and repeated focused viewings occurred on consecutive days and over the course of the 10-week curriculum so that content, vocabulary, and skills introduced in earlier weeks were revisited in following weeks.

### Intensive Support for Implementing Early Childhood Educators

Preschools participating in the science curriculum received teachers' guides containing daily scripts and 10 weeks of activities, as well as materials such as potting soil, science journals, and a stipend for perishable items used during investigations. To increase the depth and quality of implementation, teachers received ongoing professional development from an instructional coach. This training began with two two-hour orientations to familiarize teachers with the curriculum and materials. Coaches subsequently provided on-site support that included modeling examples of good teaching practice, observation, and constructive feedback, as well as general assistance with implementation. Coaches made an average of eight visits to each class during implementation. The average visit was about two hours long. Between visits, coaches provided support by telephone and e-mail.

## The Current Study

The research question investigated was: *Can a media-rich science intervention implemented by teachers in early childhood centers positively impact low-income children's talk about science with their caregivers?*

The researchers conducted a randomized experiment to test the impact of the curriculum. They randomly assigned preschool teachers in each participating center to either a treatment or comparison group. The treatment group of educators implemented the science curriculum, while the comparison group implemented a literacy curriculum of the same duration and with rich media components. The use of the literacy curriculum was to facilitate clear interpretation of findings; evaluators wanted to avoid the possibility that results could be explained by children's excitement about media use.

## Sample

Eighty classes at 47 different early childhood education centers participated in the study. Some of these centers were some part of large-scale programs like Head Start, some were run by school districts, others were small, privately run facilities, some of which were home based. More than two-thirds of the early childhood educators (69%) had a postsecondary degree.

A total of 398 children (mean age at pretest, 4 years, 9 months) participated in the study. Fifty-three percent of the children were Hispanic, and 28 percent were African American. A majority (60%) of these children spoke English at home, and a third (31%) spoke Spanish at home. Some (8%) spoke both languages at home. Over all the children in the sample, 20 different languages were spoken at home. Seventy-nine percent were low-income, as indicated by the fact that their families were eligible for a subsidy to help them pay for tuition costs.

## Outcome Measure

A key challenge in measuring outcomes of preschool curricula in science is the paucity of measures of student learning (Greenfield et al., 2009). Given the resources of this study, it was not feasible to directly measure the impacts of the science curriculum by individually assessing children. Instead, the study team relied on a single measure of caregiver reports of children's talk about science. More intensive analyses of children's science talk in the home indicate the home is a site where children's interests are expressed and, in some cases, developed with the help of caregivers (e.g., Tenenbaum, Snow, Roach, & Kurland, 2005). Our measure aligned with Strand 1 of the NRC's (2009) informal science learning guidelines and sought to capture children's excitement, interest, and motivation to learn about science topics. Caregiver reports are a less precise way to learn about children's expressed interests, but they are more readily collected than discourse data for large-scale studies such as this one.

The measure used was a single index comprised of the sum of caregivers' response to 8 items. The study team asked caregivers to report on whether or not children reported in recent weeks that they had: pretended to be a scientist or science teacher, talked about doing investigations or experiments, expressed curiosity about the natural world, expressed an interest in why things change, expressed an interest in how things work, talked about freezing and melting, talked about decay, and talked about heat and change.

## Procedure

After random assignment had been completed, the study team provided training to coaches. The training for coaches included a review of the curriculum's goals and activities, as well as guidance regarding how to support teachers in learning how to implement the curriculum. The training also included a review of a weekly log form that coaches were to complete and that provided information on teachers' implementation of the curriculum. The team supported coaches' work by providing weekly opportunities to review progress and challenges with a coach coordinator, one in New York and one in California.

After coach training had been conducted, coaches established contact with classroom teachers to set up an orientation session to the curriculum. During the orientation session, coaches reviewed the curriculum goals, activities, and strategies, as well as expectations for implementation. In addition, the coaches reviewed materials that had been shipped to teachers and that included all the digital and print content needed to implement the curriculum. In the first 4 to 5 weeks, coaches led activities that had been scheduled that day, co-led them with teachers, or observed as teachers implemented the activities. In subsequent weeks, the study team gave coaches discretion as to how many times to visit sites, though they continued to make weekly contact with sites to complete implementation logs.

Within four weeks of completing the 10-week curriculum, the study team began contacting caregivers to conduct the survey. The team made up to 20 attempts to reach parents, caregivers, or guardians who had given consent for their children to participate in the study. Response rate for the parent survey was 56 percent, which was high for parent surveys in general and for this population in particular.

## Approach to Analysis

In the study, children were nested within classrooms; classrooms were the unit of treatment assignment for the study. Because of the multilevel nature of the data, and because a significant proportion of the variance for each of the four outcomes was associated with classrooms, the study team employed hierarchical linear modeling (Raudenbush & Bryk, 2002) to estimate treatment effects. After fitting a fully unconditional model to the data to estimate variance at each level (96% at Level 1, 4% at Level 2), two models were fit: one that included only the treatment indicator as a predictor and a second that examined the impact of the parent respondent on outcomes (whether the respondent was the father or another caregiver).

Prior to modeling, the study team conducted a factor analysis and analyzed the reliability of the outcome measure. These analyses suggested a single-factor structure, and together, the items had a modest internal reliability of $\alpha = 0.75$. Since scores were skewed, the team used a bootstrap resampling method to calculate the standard deviation of parameter estimates and transformed the outcome data (using the square of the measure). Squaring the outcome measure did not produce different model results, so outcome measure for all subsequent analyses was the mean score for the 8 items, to facilitate interpretation.

Because there were missing data, the study team used data for all children in the study, namely their gender, ethnicity, and family income level, to determine whether the students' caregivers with missing data differed in any way from those for whom their caregivers did complete surveys. The results of this analysis indicated that there were no differences between the group for whom complete data were available and the group for whom data were missing on any background characteristics. The science and comparison curriculum groups were also equivalent on these background measures and a measure of early literacy.

## Results

There was a statistically significant treatment effect of the science curriculum on science talk for both models fit to the data ($p < .05$). In Model 1 (Table 1), the mean outcome score of the science curriculum children was 0.712, compared with 0.634 for the comparison group children. The magnitude of the effect was 0.30 standard deviations. There was no statistically significant relationship between child gender and scores on the outcome measure.

Table 1: Impacts of the Curriculum on Science Talk: Model Results.

| | Model 1 | Model 2 |
|---|---|---|
| Intercept | 0.634 | 0.614 |
| Treatment is Science | 0.078* (0.0335) | 0.111** (0.0332) |
| Respondent is Father | | 0.161** (0.0545) |
| Respondent is Other | | 0.000 (0.171) |
| Treatment X Father | | -0.281* (0.1199) |
| Treatment X Other | | -0.101 (0.23484) |
| $r^2$ | 0.022 | 0.052 |
| *Adjusted* $r^2$ | 0.018 | 0.014 |

* $p < .05$, ** $p < .01$

Model 2 in Table 1 presents a slightly different picture of the results and shows that when the mother was the caregiver responding to the survey, the average score for children in the science condition was 0.725 and for comparison children it was 0.614. When the father was the caregiver responding to the survey, scores were higher for both treatment and comparison groups. Moreover, when fathers are respondents, scores of children in the literacy classrooms are significantly higher than scores of children participating in the science curriculum. The data do not permit the research team to analyze why the effect is different for fathers, but this could be a topic for future research.

## Discussion and Conclusion

This study provides preliminary evidence that preschool teachers can implement a media-rich science curriculum that can impact caregiver reports of children's talk about science. Caregiver reports of children's talk reveal something of what children find interesting and memorable enough to relay about their experiences in preschool. Preschoolers' reports of recent events are not easily recalled, so the finding that children in the study talked about topics in the curriculum is especially significant.

A key limitation of the study is the outcome measure, which is only an indirect measure of children's interest in science and not a good measure of what they may have learned from the curriculum. Future studies will require additional resources to implement testing of children individually, as well as better measures of learning aligned to the curriculum's goals. Although some measures are now under development, published results of validation studies are not yet available.

One conclusion that can be drawn is that the potential of this and other curricula merit further investigation. Such studies should focus not only on the outcomes or impacts of curricula but also on the processes by which students learn from them. Studies that can identify ways that hands-on activities and interactions with video and online games complement one another would be of particular interest. Differences among learners are likely to arise, and in the context of field-based studies, so, too, will differences in preschool teachers' ability to implement the curriculum under varying levels of coaching support. In addition to understanding impacts of curricula on student learning, developing an understanding about who learns what, and under what circumstances, are important goals for learning sciences research.

## Acknowledgments

Funding for this research comes from cooperative agreements with the U.S. Department of Education (U295A005003 and U295B005003) and with the National Science Foundation (OMA-0835854). All opinions and findings expressed herein are the sole responsibility of the authors.

## References

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, J. C. (2001). Early childhood television viewing and adolescent behavior. The recontact study. *Monographs of the Society for Research in Child Development, 66*(1), 264.

Barker, T., & Torgesen, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research, 13*(1), 89-103.

Ball, S., & Bogatz, G. A. (1970). *The first year of Sesame Street: An evaluation*. Princeton, NJ: Educational Testing Service.

Bogatz, G. A., & Ball, S. (1971). *The second year of Sesame Street: A continuing evaluation*. Princeton, NJ: Educational Testing Service.

Chalufour, I., & Worth, K. (2003). *Discovering nature with young children*. St. Paul, MN: Redleaf Press.

Chambers, B., Cheung, A., Madden, N. A., Slavin, R. E., & Gifford, G. (2006). Achievement effects of embedded multimedia in a Success for All reading program. *Journal of Educational Psychology, 98*(1), 232-237.

Clark-Chiarelli, N., Gropen, J., Chalufour, I., Hoisington, C., & Eggers, P. (2009, March). *Creating a successful professional development program in science for Head Start teachers and children: Understanding the relationship between development, intervention, and evaluation.* Paper presented at the SREE Second Annual Conference, Crystal City, VA.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Reseach Journal, 45*(2), 443-494.

Crowley, K., Callanan, M. A., Jipson, J., Galco, J., Topping, K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity. *Science Education, 85*(6), 712-732.

Davidson, M. R., Fields, M. K., & Yang, J. (2009). A randomized trial study of a preschool literacy curriculum: The importance of implementation. *Journal of Research on Educational Effectiveness, 2*, 177-208.

Fisch, S. M. (2004). *Children's learning from educational television:Sesame Street and beyond*. Mahwah, NJ: L. Erlbaum Associates.

Fisch, S. M., Truglio, R. T., & Cole, C. F. (1999). The impact of Sesame Street on preschool children: A review and synthesis of 30 years' research. *Media Psychology, 1*(2), 165-190.

Foster, K. C., Erickson, G. C., Foster, D. F., Brinkman, D., & Torgesen, J. K. (1994). Computer administered instruction in phonological awareness: Evaluation of the DaisyQuest Program. *Journal of Research and Development in Education, 27* (2 ), 126-137

French, L. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Childhood Research Quarterly, 1*(1), 138-149.

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly, 19*(1), 150-158.

Gelman, R., & Lucariello, J. (2002). Role of learning in cognitive development. In S. S. Stevens & H. E. Pashler (Eds.), *Stevens' handbook of experimental psychology: Learning, motivation, and emotion* (Vol. 3). New York: Wiley.

Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the preschool program: A programmatic research agenda to improve school readiness. *Early Education and Development, 20*(2), 238-264.

Healy, J. M. (2003). Cybertots: Technology and the preschool child. In S. Olfman (Ed.), *All work and no play...: How educational reforms are harming our preschoolers* (pp. 83-111). Westport, CT: Praeger.

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development, 1,* 113-129.

Linebarger, D. L., & Walker, D. (2005). Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist, 48*(5), 624-645.

Morrow, R. W. (2005). *Sesame Street and the reform of children's television*. Baltimore, MD: Johns Hopkins Press.

National Association for the Education of Young Children (2005). *NAEYC Early Childhood Program Standards and Accreditation Criteria*. Washington, DC: Author.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits.* Washington, DC: National Academies Press.

Neuman, S. B. (1995). *Literacy in the television age: The myth of the TV effect*. Norwood, NJ: Ablex.

Pasnik, S., Strother, S., Schindel, J., Penuel, W. R., & Llorente, C. (2007). *Report to the Ready To Learn initiative: Review of research on media and young children's literacy*. New York, NY and Menlo Park, CA: Education Development Center and SRI International.

Pea, R. D., & Collins, A. (2008). Learning how to do science education: Four waves of reform. In Y. Kali, M. C. Linn & J. E. Roseman (Eds.), *Designing coherent science education* (Vol. 3-12). New York: Teachers College Press.

Prince, D. L., Grace, C., Linebarger, D. L., Atkinson, R., & Huffman, J. D. (2002). *A final report to Mississippi Educational Television*. Washington, D.C.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rule, A. C., & Guggenheim, S. (2007). A standards-based curriculum for clay science. *Journal of Geoscience Education, 55*(4), 257-266.

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly, 19*(1), 99-120.

Songer, N. B., Lee, H.-S., & Kam, R. (2002). Technology-rich inquiry science in urban classrooms: What are the barriers to inquiry pedagogy? *Journal of Research in Science Teaching, 39*(2), 128-150.

Tenenbaum, H. R., Snow, C. E., Roach, K., & Kurland, B. (2005). Talking and reading science: Longitudinal data on sex differences in mother-child conversations in low-income families. *Journal of Applied Developmental Psychology, 26*, 1-19.

Thakkar, R. R., Garrison, M. M., & Christakis, D. A. (2006). A systematic review for the effects of television viewing by infants and preschoolers. *Pediatrics, 118*, 2025-2031.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655-684.

# Scaffolding students in evaluating the credibility of evidence using a reflective web-based inquiry environment on Biotechnology

Iolie Nicolaidou, Eleni A. Kyza, Cyprus University of Technology,
P.O. Box 50329, 3603, Limassol, CYPRUS
Frederiki Terzian, Andreas Hadjichambis, Dimitris Kafouris, Cyprus Ministry of Education and Culture
Email: iolie.nicolaidou@cut.ac.cy, eleni.kyza@cut.ac.cy, freda.terzian@yahoo.com,
a.chadjihambi@cytanet.com.cy, dimitris_kafouris@yahoo.co.uk

**Abstract:** This case study investigated scaffolding to support twelve 11th grade students' collaborative construction of evidence-based explanations and their evaluation of the credibility of evidence through the utilization of a web-based reflective inquiry environment on Biotechnology. Over eleven 90-minute lessons students investigated and evaluated scientific data relating to the cultivation of genetically modified plants. The analysis of pre- and post-tests on students' conceptual understanding of Biotechnology topics and their skills in evaluating the credibility of evidence revealed learning gains and suggested that the intervention was successful. Students' written explanations in task-related artifacts and the analysis of two groups' videotaped discussions showed that the students became sensitive to credibility criteria, questioned the sources of data and correctly identified sources of low and high credibility. Students' difficulty in applying methodological criteria suggests that this criterion should be addressed in future studies.

## Introduction

A central goal of science education is to support students' evidence-based reasoning by engaging them in seeking evidence and using it to critique scientific claims. According to the U.S. National Education Standards (NRC, 1996) students should be able to apply scientific reasoning to participate in informed decision-making at the local and national level about issues that impact on their everyday life, such as the global climate change and genetic engineering. Participating in such decision-making requires an understanding of the nature of science, as well as scientific knowledge and skills relating to the interpretation and weighing of evidence, issues which are underrepresented in the current practice of teaching and learning science (Chinn & Malhotra, 2002). Reaching decisions on many socio-scientific debates that scientists and the public need to take action on requires the existence of processes such as making sense of complex and diverse data sets that are difficult to analyze and comprehend, and weighing the relevance and examining the credibility of scientific evidence. "Practices such as weighing evidence…and evaluating the potential viability of scientific claims are seen as (some of the) essential components in constructing scientific arguments (Driver, Newton & Osborne, 2000, p.288). Driver et al. (2000) noted that insufficient time is typically given to evaluative tasks beyond the interpretation of data; for instance, questions such as "what trust can we place in data?" or "are there different possible interpretations of this data?" are not frequently addressed. Indeed, often existing inquiry curricula have explicit or implicit expectations that students treat data as non-biased and do not raise concerns over the credibility of the evidence.

As Yang (2004) pointed out, given the complex issues associated with the use of evidence in ill-defined problems there is a need for more intensive investigation of students' potential to use multiple sources of evidence and to critically judge claims based on evidence. Students' engagement in solving complex socio-scientific problems provides a good outlet for having them evaluate the credibility of multiple pieces of evidence. Socio-scientific issues are real-life problems that bombard the citizens of modern society and frequently represent dilemmas, stemming from ethical aspects that need to be taken under consideration (Sadler, 2004) and from biased interpretations or presentation of the data. As Gieryn (1999), a sociologist of science, argues "what happens in nature…depends upon the chef you ask: for some, nature is a seasoning thrown in to the flavor the social meat and cultural potatoes; for others, nature is what is finally brought to the table, what gets ladled into bowls, either thick stew with chunks of social left in or thin broth after the "meat" is methodically strained out and discarded; still others never bother to pick up any nature at the market –it is social down to the bottom of the pot." (Gieryn, 1999, p. ix.)

The work presented in this paper examines 11th grade students' assessment of the credibility of evidence and describes a research-validated pedagogical approach in supporting students' attending to such issues. This work contributes to the understanding of students' capacity to evaluate the credibility of evidence, given the need to prepare future citizens in critically evaluating the credibility of evidence and the limited discussions in the literature about students' evaluation of the credibility of evidence.

## Theoretical framework

### Defining the assessment of the credibility of evidence

We draw on Driver et al. (2000)'s work to provide an operational definition of evaluating the credibility of evidence. The credibility of different sources of information can be defined as the consideration of the grounds for confidence through the use of interrogatory tools, such as the critical asking of "reports about the origin of evidence, whether the evidence is simply correlational or whether there is a plausible theoretical mechanism, whether the results are reproducible, whether they are contested, or about the authority of the scientific source" (p. 301). Put simply, the researchers referred to two main parameters that are important for the assessment of the credibility of evidence: the source of the evidence and the methodology of the construction of the evidence (Driver et al., 2000). Learners need to examine the source of the evidence and think about questions such as: Is there evident bias or not? Was a piece of evidence peer-reviewed? Who is the author of the evidence? What is the author's agenda/background? What was the source of funding for producing each piece of evidence? As far as the methodology is concerned, learners need to think about the following: Does the evidence refer to a comparison of two different groups? Is there any control of variables? Were the results replicated?

### Students' difficulties in evaluating the credibility of evidence

There are several reports in the literature on studies of credibility and how to better support students in their credibility evaluation skills; however, only a small subset of those studies come from science education, with most of them focusing on topics such as internet searches and evaluating information online. Science education research has demonstrated students' difficulties in evaluating evidence to construct evidence-based explanations at the middle school level (Glassner, Weinstock & Neuman, 2005), high school level (Sandoval & Millwood, 2005; Dawson & Venville, 2009) and undergraduate level (Lippman, Amurao & Pellegrino, 2008). One of the most important and common difficulty students have in relation to evaluating evidence refers to an uncertainty as to what constitutes convincing evidence or valid evidence (Driver et al., 2000). Furthermore, previous research showed that one of the important reasons students faced difficulties in evaluating the credibility of evidence was the fact that they lacked the criteria they needed to be able to evaluate evidence (Wu & Hsieh, 2006).

At the middle school level, Pluta, Buckland, Chinn, Duncan and Duschl (2008) documented a widespread difficulty that the participants in their study, 724 US middle school students, faced with regard to the effective use of reasons and evidence. Students typically found it difficult to make judgments about the relative strength of evidence, and rather tended to treat all evidence as equally strong. They also had difficulty in understanding the need to provide justifications that were more elaborated than just mentioning the evidence that provided support. Seethaler and Linn (2004) examined 190 $8^{th}$ grade students' reasoning about tradeoffs in the context of a technology enhanced curriculum about genetically modified food and found that even though students were able to provide evidence both for and against their positions, they were less explicit about how they weighed those tradeoffs.

At the high school level, Kolsto (2001) interviewed 22 $10^{th}$ grade students on their views on the trustworthiness of knowledge claims, arguments and opinions on a socio-scientific issue. All students expressed problems and uncertainty when trying to sort out who to trust and what to believe. As a result, some claims were not questioned at all. Moreover, most students showed very little interest in empirical evidence underpinning knowledge claims and only a few expressed a positive interest in methodological aspects to assess the credibility of evidence presented to them.

Even at the college level, students sometimes had difficulty assessing the credibility of evidence in scientific arguments. Lippman et al. (2008) investigated 98 college students' understanding of evidence use in quality scientific arguments, by measuring students' ability to compare and analyze the quality of arguments and the use of evidence in essays. Their findings revealed that students inconsistently applied the epistemological criteria of empirical data serving as evidence in scientific arguments. Their insensitivity to blatantly inaccurate descriptions of evidence suggests students lack the deep understanding needed to assess the quality of evidence. They made reasonable, but not always ideal selections for the strong and weak evidence and focused on superficial aspects, such as ease of comprehension instead of the need for relevant empirical and disconfirming evidence.

It seems that training can support students in evaluating evidence. With respect to evaluating evidence, Sanchez, Wiley and Goldman (2006) reported that college students had a fragile understanding of the credibility of sources of evidence in internet sites and very few could verbalize it or use it to justify their evaluations of credibility. The researchers identified four key areas in which college students needed support: considering the source of the information, considering the evidence that was presented, thinking about how the evidence fit into an explanation of the phenomena and evaluating the information with respect to prior knowledge. In their study with 60 undergraduates who were randomly divided into two groups, one group of students received a short training course on criteria that could be used to evaluate internet sites such as: who the author of the information was, how reliable a site was, how well the site explained the information, identifying relevant information about the author, such as motivation, examining whether the information was consistent with other reliable sources and whether it was based on scientific evidence. Results indicated that college students were able to learn to use critical evaluation skills during short-term training and were better able to identify relevant from irrelevant sources.

The present study examined students' difficulties in evaluating the credibility of evidence during a decision-making process. The following research questions were addressed: How do 11th grade students evaluate the credibility of different sources of data when trying to reach a decision on a complex socio-scientific issue? What evidence do they use and how do they use it? How can this process be scaffolded?

## Methodology

### Participants

Two 11th grade intact classes in two public high schools in Cyprus participated in this study. The first class consisted of twelve students who took the course "Humans and Health" as an elective. The students of the class, in this case average students (mean GPA=14.7/20), were randomly assigned in four heterogeneous groups of three students each. The second class served as a control group and consisted of 13 students who took the course "Human and Health" as an elective in a different school.

The enacting teacher was a Biology teacher with a Bachelor's degree in Biology and 17 years of teaching experience. She is one of the authors and also a member of the group which designed the learning environment. The first author was present in the class during all lessons of the enactment to provide support with regard to the use of the web-based learning environment.

### The intervention: The Biotechnology and Genetically Modified plants learning environment

A problem-based learning environment on the topic of Genetically Modified Organisms (GMOs) was developed using the web-based platform of STOCHASMOS (Kyza & Constantinou, 2007). STOCHASMOS is a learning and teaching platform for supporting students' scientific reasoning through scientifically authentic investigations with an embedded authoring tool that can be used by teachers or instructional designers. The driving question that students were asked to answer was whether they would allow the cultivation of genetically modified (GM) plants in their country. Students were asked to make sense of real and diverse data sets mostly coming from scientific publications that were comparing GM and conventional plants from the perspective of the environment, economy and health. Data took multiple formats and included sources of low, average and high credibility. An example of high credibility data was the result of experimental research (e.g. research that entailed the comparison of an experimental and a control group and was published in a peer-reviewed scientific journal). An example of low credibility data was an opinion-based article, which reflected the views of the author, was not accompanied by data and did not substantiate claims.

The activity sequence consisted of eleven 90-minute lessons. The first three lessons consisted of hands-on experiments with DNA extraction, while for the next eight lessons students worked collaboratively through the STOCHASMOS learning environment of Biotechnology. A 40-min section in the fourth lesson was devoted to having students evaluate the credibility of two data sources in a context unrelated to Biotechnology. The aim of this activity was to reveal the criteria that students were intuitively using to evaluate credibility and to guide them into identifying the following criteria that were part of the scaffolding within STOCHASMOS: "author background", "type of publication", "funding" and "methodological criteria". In lessons 5 to 10, students evaluated data, captured evidence and went through the process of interpreting and explaining the data organized in the STOCHASMOS WorkSpace templates to develop evidence-based explanations about whether they would allow the cultivation of GM plants in their country. Students were prompted to evaluate the credibility of evidence throughout their investigation. The lessons concluded with the groups' in-class presentations of their work.

### Scaffolding through a reflective web-based inquiry environment

For the purposes of this research study, to "evaluate the credibility of evidence" means to assess the credibility and value of data that has been determined to be relevant for an inquiry-based investigation. Scaffolding was both human-provided and technology-supported. The scaffolding provided by the teacher included, among others, explicit instruction on how evidence and explanations are evaluated, based on given criteria. The scaffolding provided by peers included peer-evaluation of explanations, incorporated through the sharing work capability of the web-based platform used and the use of the chat-tool.

Technology-supported scaffolding was provided through the STOCHASMOS platform. The learning environment included the use of templates, which structured the students' task in creating evidence-based explanations for the impact of GM plants on economy, environment and health. The templates included passive prompts encouraging students to identify the source of the evidence and rate its credibility using a scale of 1 (low credibility) to 5 (high credibility). Drawing from the work of Driver et al. (2000), the main criteria for the evaluation of evidence in the context of this study were twofold and referred to: a) the source of the evidence and b) the methodology of the construction of the evidence. The source of the evidence referred to: author's background, type of publication and funding. Methodology was limited to having students examine whether the evidence referred to a comparison of two different groups, e.g. GM plants and conventional plants in this case.

### Data collection and Analysis

When designing this case study we sought to collect diverse data that would maximize triangulation and increase the validity of the conclusions. Thus we collected both outcome and process data, at the individual students' level as well as at the collaborative, group level. Data sources included pre- and post-tests on students' conceptual understanding of Biotechnology concepts and the credibility of evidence and all groups' videotaped discussions throughout the enactment. The tests that were administered to the enacting group were also administered to the control group twice over a period of three months, which was the approximate duration of the enactment, to control for the testing effect. The control group students did not take any Biotechnology related lessons.

The conceptual understanding individual test consisted of ten multiple choice questions and seven open-ended questions on topics related to Biotechnology. A scoring rubric was developed for scoring students' answers by the first two authors. Interrater reliability was computed (r=0.79). A test examining individual students' evaluation of the credibility of sources was administered prior and after the enactment. The test consisted of two open-ended tasks. The first task asked students to evaluate the credibility of a given source. The rationale behind this task was to assess the criteria that students were intuitively using to evaluate credibility. In the second task, students had to compare the credibility of two sources to decide which cell phone model to buy. The rationale behind this task was to assess students' skill in differentiating between a credible (results of a study that used random sampling, control of variables, comparison of two groups, and was published in an independent magazine) from a non-credible publication (advertisement publication indicating source bias as it came from a company that sold the product and had economic interest). In both tasks students had to justify their answer. A scoring rubric was developed for scoring students' answers by the first two authors. Interrater reliability was computed (r=0.83). Students' responses were also examined qualitatively, using phenomenographic analysis methods to extract categories from students' answers.

All lessons during the enactment were videotaped. Two of the four groups of students, Group 1 (mean GPA=14.9/20) and Group 3 (mean GPA=16.2/20), were selected for in-depth analysis. The first researcher viewed all videotaped lessons twice to isolate meaningful episodes that related to issues of assessing the credibility of evidence; a list of relevant keywords (e.g. evidence, credibility, believable, source, author, bias, results, comparison, type of publication, funding, graph, internet, scientist, journal, and the rating that students used to evaluate credibility) was identified a priori to help guide this process. A total of 18 episodes were identified for Group 1, ranging from 1 to 10 minutes each (total= 61 minutes) and 14 episodes for Group 3, ranging from 1 to 4 minutes each (total=29 minutes). Three additional episodes referred to the teacher talking about credibility in whole class discussion activities. The episodes were transcribed in the local language. They were then analyzed with respect to the four criteria that students were asked to evaluate: author background, funding, type of publication and comparison between two groups, to identify students' difficulties and describe the development of students' thinking with regard to the credibility of evidence over time. We also analyzed the data to examine whether discussions were scaffolded and by whom or what (e.g. took place while students were working within the STOCHASMOS environment or took place after prompting by the teacher).

## Results

### Effectiveness of intervention

To assess students' learning gains, we compared their pre- and post-tests using the Wilcoxon signed ranks non-parametric test because of the small sample size. The analysis indicated a statistically significant difference $z(12)=-2.91$, $p<.01$, and an effect size of 0.59. The enacting students' performance increased from a mean of 16.58 (SD=5.02) in the pre-test to a mean of 24.17 (SD=6.13) in the post-test. The maximum possible score of the test was 38. The control group students' performance increased from a mean of 13.5 (SD=6.13) in the pre-test to a mean of 14.5 (SD=5.23) in the post-test. However, the analysis using the Wilcoxon signed ranks non-parametric test did not indicate a statistically significant difference for the control group students $z(12)=-1.27$, $p>.05$.

## Students' evaluation of the credibility of sources pre- and post-enactment

### Results from the analysis of individual students

The students' performance increased from a mean of 4.5 (SD=2.24) in the pre-test to a mean of 8.42 (SD=3.40) in the post-test (maximum possible score=12). To assess students' gains with regard to their ability to evaluate the credibility of sources, we compared their pre- and post-tests using the Wilcoxon signed ranks non-parametric test. The analysis indicated a statistically significant difference, $z(12)=-3.09$, $p<.01$, and an effect size of 0.63. The control group students' performance increased from a mean of 4.75 (SD=2.01) in the pre-test to a mean of 5.20 (SD=3.36) in the post-test (maximum possible score=12). The Wilcoxon signed ranks non-parametric test analysis did not indicate a statistically significant difference for the control group students $z(12)=-1.14$, $p>.05$.

How did students' understanding of the credibility criteria change? In addition to the quantitative analysis of students' skills in evaluating the credibility of sources, a phenomenographic analysis of their answers was also conducted. As a result, students' answers were categorized in levels which are presented in Tables 1 and 2. The analysis of the first task, which asked students to evaluate whether a publication was credible or not, showed that students' answers shifted from opinion-based answers (based on merely copying information from the text), to more sophisticated answers (which identified one or more criteria for evaluating credibility). Table 1 shows the number and percentage of students' answers in each level for the pre-test and post-test (n=12), as well as illustrative examples of the scoring.

Table 1:Task 1 "How did students evaluate the credibility of a given source?"

| Category | Illustrative examples | Pre-test | Post-test |
|---|---|---|---|
| Level 1: Opinion-based/copying information from text | "I have the opinion that a person can only fight addiction with her own will (…) Therefore, I don't believe this publication". | 3 (25%) | 0 |
| Level 2: Source provided details or numbers or results or statistics | "It (the research study) is not convincing because it does not mention details (…) On the other hand it mentions statistical results". | 9 (75%) | 0 |
| Level 3: Based on one criterion (author, funding, source type, or methodology) | "The research study was conducted by the Acupuncture Center and it was published by the same center. Therefore they had interest in having positive results for advertisement purposes". | 0 | 8 (67%) |
| Level 4: Based on at least two of the criteria | "The duration of the research study was adequate, the participants were men and women of various ages, the publication was from a university, the author was a doctor, a general practitioner. I don't think Dr. X. had any financial interest, because it's his duty to help people so I believe in this research study". | 0 | 4 (33%) |

The analysis of the pre-test showed that most students (75%) initially thought that if a source provided details and numbers, or results and statistics it meant that it was credible (Table 1). The rest of the students (25%) provided an answer that relied on either their own opinion or verbatim information from the text. However, none of the answers on the post-test fell into the pre-test categories. Students based their answers on the application of one (67%), or two and more credibility criteria (33%). This suggests that by the end of the intervention students had appropriated these criteria.

The second task asked students to choose which one of two given publications was the most credible. Table 2 shows the number and percentage of students' answers in each level for the pre-test and post-test for task 2 (n=12), as well as illustrative examples of the scoring in each level.

Table 2: Task 2 "How did students compare the credibility of two given sources?"

| Category | Illustrative examples | Pre-test | Post-test |
|---|---|---|---|
| Level 1: No justification or opinion-based or copying information from text | "I think that the most convincing is the second publication because it seems more compelling". | 4 (33%) | 1 (8%) |
| Level 2: Source provided results | "It described the advantages of the cell phone in detail and provided specific results" | 1 (8%) | 0 |
| Level 3: Identified "source bias" only | It was "an independent research study (…) published by an independent company". | 1 (8%) | 4 (33%) |

| | | | |
|---|---|---|---|
| Level 4: Identified "comparison" only | "It presents a cell phone and compares it to another cell phone". | 5 (42%) | 0 |
| Level 5: Identified both "source bias" and "comparison" | "The publication was conducted by an independent company (…) and there was a comparison of two cell phones and the participants of the survey were chosen randomly". | 1 (8%) | 7 (58%) |

In the pre-test several students identified the fact that a comparison was conducted (42%), one student was able to indentify the bias of the source (8%), and another student identified both criteria (8%). In contrast, in the post-test, the vast majority of students (91%) correctly identified either "source bias" or both criteria.

## Students' collaborative shift from opinion-based explanations to criteria-based explanations

An analysis of the 40-min section in the fourth lesson, which was devoted to having students evaluate the credibility of two data sources in a context unrelated to Biotechnology, was conducted to examine the criteria that students were intuitively using to evaluate credibility prior to working in STOCHASMOS. This showed that students were capable of identifying the "author background" criterion unassisted, but needed the teachers' support to identify other criteria for credibility, such as "type of publication", "funding" and "methodological criteria".

An analysis of the total of 35 episodes revealed that the vast majority of students' discussions on credibility of evidence (24 episodes) were prompted and scaffolded as they took place while students were working with the STOCHASMOS templates. Only two episodes took place spontaneously. The teacher prompted students to consider the evaluation of the credibility of their sources in two of the episodes, while another two episodes referred to teacher-initiated whole-class discussions. In the remaining five episodes students attempted to verify their sources. The analysis of students' videotaped discussions revealed several themes and illustrated the shift of the students of the two selected groups (n=6) from level 2 in the pre-test to levels 3 and 4 in the post-test.

### Students became critical of and questioned the credibility of their sources

Baseline pre-test data showed that students sometimes even accepted advertisements as credible pieces of data and did not question the source of information. During the investigation, students became very critical of the sources they were basing their decision on. When Group1 students compared the expenses and income of GM to conventional maize they found the source non-credible. They expressed their concern that they were basing their conclusions about the effect of GM plants on economy on evidence that they did not find credible and they were puzzled by this fact. The group had some insecurity about their work and asked the teacher whether it was acceptable to make some observations and interpretations of data and reach some conclusions but then cite and evaluate their source only to find that they did not consider it a credible one. The same issue was troubling for students of Group 3.

Throughout the STOCHASMOS enactment and through their interaction with the learning materials and with their peers, students formed opinions about source credibility. These two groups' discussions about credibility focused on the need to evaluate all data sources and students gradually developed a critical stance toward the sources they were examining during their inquiry, which was evident in lessons 4 to 11.

### Students showed distrust to sources of average credibility and were sensitive to source bias of "funding" and "type of publication"

As expected, sources of average credibility stimulated the most discussion among students during their evaluation. Students correctly showed distrust to those sources and were troubled as to what they should believe and what not. During the fourth lesson, Group 1 focused on the impact of GM maize on economy and examined financial data. The discussion showed the distrust of this group to the source, mainly because the study had been conducted by a financial advisor in a company that provided advice to biotechnology companies and it was published on the website of the company, receiving funding by the same company. In general, students showed distrust in anything that was published on the internet. Student names are pseudonyms.

*Group 1: Lesson 4: min 2:24-11:00*

*Angelina:* *"…to evaluate the credibility of the source. "Author background", Brook is the financial advisor of PG economics, which provides advice to biotechnology companies. I don't believe him because everything is done by the company. And they make biotechnology products. And in the comparisons, it says here that the GM is less than the conventional (talking about the expenses of cultivation) and he has the company for biotechnology so (...)"*
*…I am not convinced. It's all for their benefit!"*

*Kate:* *Now let's look at funding. Since it was funded by the same company it's not credible. (Should we rate it with) One (1) or two (2)?*
*Angelina:* *One (1)!*
*Kate:* *It's not convincing at all (...) because they published it on the website of their own company.*

Similarly, in their attempt to evaluate the credibility of the same source, Group 3 students also identified funding as a potential problem and had the same strong idea that anything that was published on the internet was necessarily non-credible. However, it seemed that students gradually overcame that misconception. During lesson 8, Group 1 evaluated two sources on the impact of GM maize on people's health. Even though initially the group tended to believe that whatever was published on the internet was automatically non-credible, in this case they correctly identified that even though the publication was online, the source citation indicated that it was actually a journal publication.

*Group 1: Lesson 8*
*Rania:* *Wait. Yes, but (reading the title of the source), issue 29, p.26-28. Why would it then say publication on the internet? (laughing) Here I go again with my questions!*
*Kate:* *Wait. Here it's not convincing because it was on the internet.*
*Rania:* *Why does it say publication of the journal?*
*Aggela:* *Ok, they took a part and then put it on the internet.*
*Kate:* *Yes. So it* ***is*** *convincing (her emphasis).*

### Students easily identified journal publications as high credibility sources
Both Groups 1 and 3 had no difficulty evaluating a source that appeared in a peer-reviewed scientific journal, received funding from independent research organizations, was conducted by independent researchers or professors and included reported methodology that included a comparison of GM and conventional plants as a source of high credibility. Both groups rated all criteria with the highest possible score for credibility, which was 5 out of 5.

### Students had difficulty in applying methodological criteria
Students faced some difficulty in understanding what the methodological criterion of "the existence of a comparison between two groups" meant and needed support for its application. Students did not seem to understand that the criterion "existence of a comparison between two groups" referred to having an experimental and a control group, even after four lessons of applying the credibility criteria during the investigation.

## Discussion
The evidence suggests that the intervention was successful at increasing these students' conceptual understanding of Biotechnology-related concepts and at increasing their skills in evaluating the credibility of sources. Through their interaction with the learning environment, their teacher and their peers, students became critical of the sources of evidence. They paid attention to citing their sources and evaluating them when they were prompted to do so through the STOCHASMOS templates. They also paid attention to referring to their sources and commenting on their credibility during their in-class presentations of their work. Contrary to previous research, which documented middle school level U.S. students' difficulties in the effective use of reasons and evidence and stated their tendency to treat all evidence as equally strong (Pluta el al., 2008), the students in the present study had no difficulty in making judgments about the relative strength of evidence. This was evident by the finding that they correctly identified sources of low, average and high credibility. Students could also differentiate between weaker and stronger sources with regard to their credibility, as was evident by the analysis of their post-tests. This finding suggests that the intervention was successful with regard to supporting students' skills in evaluating the credibility of sources.

The fact that the majority of students' discussions regarding credibility criteria did not happen spontaneously but were, rather, a result of scaffolding showed students' need for support. The results of this study showed that a reflective web-based inquiry environment that prompted students to evaluate the credibility of their sources, combined with the use of given credibility criteria seemed to have provided an effective scaffolding mechanism toward the goal of making students develop a critical stance towards evidence. An epistemological shift was evident as students realized that there were two levels in evaluating the credibility of evidence: first an interpretation of data was necessary to make sense of the data that qualified as evidence and then as a second step, the credibility of the evidence could be assessed. They learned to question the evidence depending on the source it came from and at the same time they were puzzled and had reflective discussions about what to believe and what not, which would not have happened had they not been asked to engage in the task of evaluating the credibility of evidence they included in their investigation.

## Implications for practice

This study demonstrated the importance, necessity and effectiveness of a reflective, web-based inquiry environment as a scaffolding mechanism to support students' evaluation of the credibility of evidence. An inquiry environment that is designed to support students in their credibility evaluation skills should include sources of data of varying degrees of credibility (e.g. low, average and high credibility) to provide students with opportunities to use critical judgment to accept and reject evidence. It was also shown that the design of interventions to foster students' credibility evaluation skills can be based on the use of specific criteria. Criteria can be either provided a priori or can be derived by students. Depending on the age, grade, and achievement level of the students, credibility criteria can be either simple and focus on two major parameters, such as "source" and "methodology" or can be advanced and sophisticated and cover a greater number of aspects and parameters of credibility. The level of complexity has to be decided in advance as it directly affects the amount of information that students need to have access to in order to be able to make an informed decision regarding how believable each credibility parameter is for them. Students in this study had difficulty in focusing on methodological issues of the studies they evaluated. It is important to address methodological criteria, which are among the most important ones, in relation to the evaluation of the credibility of evidence in future studies.

## References


Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically Authentic Inquiry in Schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*, 175-218.

Dawson, V. & Venville, G. J. (2009). High-school Students' Informal Reasoning and Argumentation about Biotechnology: An indicator of scientific literacy? *International Journal of Science Education* 31 (11), 1421–1445.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the Norms of Scientific Argumentation in Classrooms. *Science Education*, 84(3), 287-312.

Gieryn, T. F. (1999). *Cultural Boundaries of Science: Credibility on the Line* (First ed.). Chicago: University Of Chicago Press.

Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology* (75), 105-118.

Kolsto, S. D. (2001). "To trust or not to trust,…"- pupils' ways of judging information encountered in a socio-scientific issue. *International Journal of Science Education*, 23 (9), p. 877-901.

Kyza, E. A., & Constantinou, C. P. (2007). *STOCHASMOS: a web-based platform for reflective, inquiry-based teaching and learning* [software]. Cyprus: Learning in Science Group.

Lippman, J. P., Amurao, F. K., & Pellegrino, J. W. (2008). Undergraduate Cognitive Psychology Students' Evaluations of Scientific Arguments in a Contrasting-Essays Assignment. Paper presented at the International Perspectives in the Learning Sciences: Cre8ing a Learning World, Utrecht, The Netherlands.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

Pluta, W. J., Buckland, L. A., Chinn, A. C., Duncan, R. G., & Duschl, R. A. (2008). Learning to Evaluate Scientific Models. Paper presented at the International Perspectives in the Learning Sciences: Cre8ing a Learning World, Utrecht, The Netherlands.

Sanchez, C. A., Wiley, J., & Goldman, S. R. (2006). Teaching students to evaluate source reliability during internet research tasks. Paper presented at the ICLS.

Sandoval, W. A., & Millwood, K. A. (2005). The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cognition and Instruction*, 23(1), 23-55.

Seethaler, S., & Linn, M. (2004). Genetically modified food in perspective: an inquiry-based curriculum to help middle school students make sense of tradeoffs. *International Journal of Science Education*, 26(14), 1765-1785.

Wu, H. K., & Hsieh, C. E. (2006). Developing Sixth Graders' Inquiry Skills to Construct Explanations in Inquiry-based Learning Environments. *International Journal of Science Education,* 28(11), 1289-1313.

Yang, F. Y. (2004). Exploring high school students' use of theory and evidence in an everyday context: the role of scientific thinking in environmental science decision-making. *International Journal of Science Education*, 26(11), 1345-1364.


## Acknowledgments


This work was funded by the 'Science in Society' Initiative of the European Commision's Seventh Framework Research Programme (FP7), under the grant CoREFLECT (217792) awarded to the Cyprus University of Technology. Opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the funding agency. For more information about the CoReflect project please visit http://www.coreflect.org

# Arguing with Peers: Examining Two Kinds of Discourse and Their Cognitive Benefits

David Shaenfield, Teachers College – Columbia University
525 W. 120th St., New York, NY, 10027, United States
des2020@columbia.edu

**Abstract:** This study examines the extent to which meta-level regulation of argumentive discourse goals and strategies relates to improvement in argumentive discourse skill. A 7-month intervention was designed to provide dense experience in argumentive discourse and to promote meta-level regulation of discourse. Pairs of academically disadvantaged eighth graders conducted electronic dialogs with opposing pairs on a series of social topics. Analysis of intra-dyad discussion over the course of the intervention showed that participants producing a high proportion of meta-level utterances were more likely to show improvement in argumentive skill. This finding suggests that enhanced meta-level awareness of the strategies and goals of argument, along with rich engagement in argumentation, promote skill development.

## Introduction

The idea that students profit from collaborative learning continues to gain popularity, and some, although not extensive, empirical evidence exists to support it. The mechanisms involved, however, have not been well established. I seek to learn more about these mechanisms in the present work. The domain in which I conduct the work, argumentation, is one based in collaborative social interaction. Indeed, this social exchange, as suggested elsewhere (Kuhn, Goh, Iordanou & Shaenfield, 2008, Sampson & Clark, 2009), serves as a promising path to development of the individual expository argument skills that figure prominently in academic contexts.

The present work includes an additional layer in this collaboration by asking two peers who share the same position on a social issue to collaborate in argumentive discourse with a series of opposing pairs who hold an opposing position. I focus on the discourse between the agreeing pairs as they engage in the joint task of formulating their contributions to electronic discourse with a series of opposing pairs. In a preliminary study that provides a foundation for the present work and employed very similar methods, we compared a condition involving collaborating pairs with a condition in which individuals worked alone in formulating and implementing the same kind of electronic arguments with an opposing peer. When subsequently assessed individually at the end of a several-month period, those in the collaborating condition showed greater increase in argumentation skill (as indexed by frequency of use of direct counterargument, the same measure employed in the present work). I therefore seek in the present study to investigate what occurs in the collaborative discourse of same-side pairs that is beneficial to their individual skill development.

The hypothesis that guided the present study is that this benefit is metacognitive in nature. Meta-level regulation lies at the end of a meta-level developmental trajectory beginning with a young child's awareness of mental functions (Kuhn, 2000; Wellman, 1992). In the argumentive context, the focus is on metastrategic knowing – specifically knowing about how to effectively engage in an argument. This metastrategic competence includes knowledge of the goals of argumentation, including attention to the opponent's position and claims, and knowledge of different discourse strategies. Awareness is a first step toward meta-level competence. A high level of such competence is reflected in metastrategic evaluation, which enables the speaker to monitor the effectiveness of strategy use. Awareness and evaluation are both necessary along a path of meta-level development toward true meta-level regulation.

Meta-level regulation in the argumentive context involves the speaker's ability to control strategy use during argumentive discourse. Speakers at this level are able to choose and sequence different strategies effectively, based on their knowledge of the utility of the strategies, while understanding how the strategies lead to the dual goals of argumentation, identified by Walton (1989) as obtaining commitments from the opposing partner to support one's own arguments and critiquing the opposing partner's arguments to weaken their force. In addition, regulation allows for continuous adjustments in response to the other speaker's discourse strategies. This metastrategic effort reflects knowledge and control over a sequence of strategies involving counterargument and rebuttal (counterargument of a counterargument).

Three techniques were employed in the present work to support students' awareness of and reflection on their argumentive discourse. These were thought to be key in enhancing both their understanding of the goals of argumentation and their skill in implementing these goals. First, students worked in pairs. Previous research demonstrates the benefit of collaborative planning and implementation in the development of young adolescents' argument skills (Kuhn & Udell, 2003; Shaenfield & Moore, 2009; Sampson & Clark, 2009).

A primary benefit of collaborative planning is making knowledge explicit. Learning is shown to be enhanced when learners articulate the reasoning behind problem-solving behaviors (Chi & Van Lehn, 1991). In the context of planning argument discourse strategies, collaboration provides the opportunity for articulation of meta-level regulatory strategies. It was hypothesized that two students with the same goal and having to deliberate regarding each argumentive move would not only generate more effective contributions to the discourse but would also make the collaborative pair more aware and reflective regarding this discourse.

Second, following the work of Felton (2004), an explicit reflective activity was added. Some of the students' own dialog transcripts were made available as the basis for this activity. Third, students conducted their dialogs via instant-messaging (IM) software. In addition to capitalizing on teens' familiarity with the medium, this technique has the benefit of providing an immediately available and permanent record of the discourse, as a basis for reflection. In striking contrast, in real-time verbal discourse, the contents of each contribution to the dialog immediately disappear as soon as they are spoken.

A number of previous studies (Anderson, Nguyen-Jahiel, McNurlen, Archodidou, Kim, Reznitskaya, et al., 2001; Chinn, 2006; DeFuccio, Kuhn, Udell, & Callender, 2009; Felton, 2004; Felton & Kuhn, 2001; Kuhn et al., 2008; Kuhn, Shaw, & Felton, 1997; Kuhn & Udell, 2003; Nussbaum, 2003, Nussbaum & Kardash, 2005; Nussbaum & Sinatra, 2003; Reznitskaya, Anderson, McNurlen, Nguyen-Jahiel, Archodidou, & Kim, 2001; Schwarz, Newman, & Biezuner, 2000; Udell, 2007) have reported gains in argumentation skill as a result of engagement and practice.

Previous research (DeFuccio et al., 2009; Felton, 2004; Felton & Kuhn, 2001; Kuhn et al., 2008; Kuhn & Udell, 2003) has shown that young adolescent arguers tend to limit themselves largely to discourse strategies addressed to articulation of their own positions. Over time, a shift is seen toward greater attention to the opponent's claims and, eventually, efforts to address and weaken them. As this occurs, a shift in the awareness of the goals of argumentation may also be developing. Growth of the meta-strategic aspect of argumentation, alongside skill development, may be critical for progress. Previous studies of argument skill development have not explicitly addressed the relationship between this meta-level development and skill development itself. To further understand the relationship between developing meta-level regulation and developing dialogic argumentation skills, the intra-dyad discussion between same-side partners was recorded on audiotape for subsequent analysis.

## Method

The research employs a two-group repeated measures design using an untreated control group to compare the effects of the intervention on argument discourse skill. The entire study took place in repeating phases over seven months.

### Participants

Participants were the entire eighth-grade student body of 49 students not assigned to individualized education programs (IEP), drawn from three classrooms, in an urban public middle school. The school primarily serves African American and Latino families from a lower-income community; 80% of students are eligible for free lunch and 14% are eligible for reduced-price lunch. The city school system classifies the student body of the school as academically disadvantaged, with student achievement is several years below grade level. The city education website lists the student body ethnicity ratios as 54% Black or African-American, 40% Hispanic or Latino, 4% White, 1% Asian or Pacific-Islander and 1% Native American. One class of 26 students (17 females and nine males), participated in the intervention, while the remaining 23 (10 females and 13 males), drawn from two classrooms, served as a control group.

### Procedure

#### Pretest Assessment

The purpose of this phase was to assess all 48 students' individual initial skill levels in dialogic argument, for comparison at final assessment. Students' opinions regarding capital punishment (CP) were assessed individually by questionnaire. This assessment allowed students to be classified as holding a pro or con position on CP. Each student was then paired with a classmate holding an opposing (pro or con) view on CP. Each pair conducted a dialog on the CP topic mediated by iChat (instant messenger chat software) installed on laptop computers. The specific instruction was to attempt to convince the other opponent that his or her opinion was the better one. Pairs of pro and con students were positioned to face opposite sides of the classroom to reduce the possibility of verbal or eye contact. The software saved a transcript of the dialog for analysis. Sessions were approximately 25 minutes in length.

## Initial Intervention

The initial phase of the intervention took place over the course of 11 successive 45-minute class periods, twice per week. On the days when the experimental group was participating in the intervention, the control group attended their regular social studies class. The experimental group's first topic was whether children should be required to attend a town school or whether it's permissible for parents to teach them at home if they wish.

*1. Individual position and assessment (session 1).* Students' positions regarding the first topic were assessed individually by questionnaire, allowing students to be classified by position on the topic.

*2. Dyadic electronic arguments with opposing-view pairs (sessions 2-7).* Students were paired with a classmate holding the same view (home okay or school mandatory). An attempt to form the greatest number of same-gender dyads was made but not possible to implement completely due to the uneven gender make-up of the intervention group. However, the majority of pairs were same-gender. The two members of a pair remained paired throughout the topic. The paired students collaborated in engaging in a series of e-dialogs with other dyads holding the opposing view. In one case, to accommodate an uneven number of students, a trio was formed. Dyads were instructed and reminded to collaborate with one another in constructing their input and, once in agreement, to take turns typing it on the laptop. An audio tape recorder continuously recorded the intra-dyad discussion. Sessions averaged 25 minutes in length.

Dyads engaged a different opposing dyad at each session. During the dialog period, several adults circulated to address any technical difficulties and to remind partners to collaborate. Dyads were also provided with a series of "reflection sheets" to promote their representing and reflecting on their own and the opposing pair's arguments. The "Other Argument" sheet asked the students to determine what the other side's main argument was, their dyad's "comeback" (counterargument) to that argument, and how the dyad could improve this comeback. The "Own Argument" sheet asked the students to identify one of their main arguments, the other side's counterargument to that argument, the dyad's comeback (rebuttal) to the counterargument, and to suggest how that comeback could be improved.

*3. Preparation for final, class-level e-argument (sessions 8 & 9).* Session 8 marked the beginning of a culminating activity with respect to the topic. Students worked within their same-side teams, preparing for what was to be a final "showdown" debate at session 11. Within each of the same-side teams, each of the dyads that had worked collaboratively to this point were divided, one assigned to an "Other Argument" (OTH) team and the other to an "Own Argument" (OWN) team. Each preparation team had an adult coach who facilitated the group process. The materials available to each team were as follows: The "Own Argument" teams had copies of the "own argument" reflection sheets that had been produced by each team. The "Other Argument" teams had copies of the "other argument" reflection sheets that had been produced by each team. Coaches had transcripts of all dialogs on hand for reference if requested by the students.

The OWN team was told that their goal in preparing for the showdown was to be familiar with the possible counterarguments the opposition could assert and prepare rebuttals to use in the showdown. The team created a set of "own argument – counter – rebuttal" sequences that were recorded onto the OWN argument sheet. To do this, first they selected the three most convincing arguments for their position. The reflection sheets were provided as references to complete the OWN argument sheet counter and rebuttal sections. The OTH team was told that their task was to be familiar with effective counterarguments to use when faced with other arguments advanced by the opponents in favor of their position. The team created a set of "other argument – counter" sequences that were recorded onto the OTH argument sheet. During this preparation session, these team members identified the three most damaging arguments raised by the opposition and, using the provided or newly generated materials, decided on their counterarguments.

*4. The Showdown (session 10).* Students on each side were divided into two teams of 6-7 members – Team A and B based on the recommendations of the coaches during the preparation phase, so as to create two teams of overall comparable skill. The division was done such that each team contained members who in the preparation phase specialized in generating the best "own arguments" for their position, as well as members who had specialized in countering the "other arguments" of the opponents.

During the first half of the showdown, Pro Team A debated with Con Team A. At half time, a team change took place and B Teams from each side continued the showdown. The showdown consisted of a single electronic dialog between the two sides lasting approximately 30 minutes. All members of the team collaborated and came to an agreement on the text to be sent to the opposition.

*5. Judging and feedback (session 11).* The electronic dialog produced in the showdown was represented in an argument map prepared by the researchers. All the arguments, as well as the corresponding counterarguments and rebuttals, were represented and connected by lines to show their interrelation. Different colors were used to label contributions as effective, ineffective, or neutral argumentive moves. A point system was also applied, which made it possible to declare a winning side. The argument map as well as the graph representing the outcome was presented to the students for their examination the week following the showdown.

### Subsequent Intervention Phases

The intervention was repeated during the second and again during the third trimester of the school year, with new topics. The second topic addressed whether disruptive students should be expelled from school or allowed to remain. The third topic, linked to the social studies curriculum, addressed whether the United States should offer military intervention to a small country requesting it. Due to time limitations created by school scheduling conflicts, the second and third interventions included only five dialogic sessions instead of six.

### Posttest Assessment

The pretest assessment involving individual dialogs on CP was repeated for all students. In the intervention group, one student participating in the initial assessment left the class after the first topic and was not included in the posttest assessment. Another student who joined the intervention classroom during the second topic was also not included in the posttest assessment.

## Results

### Overall effect of the intervention

Dialogs were coded based on the argument discourse coding scheme developed and used in previous research (Felton, 2004). The main indication of improvement was the increase in counter-critique strategy moves – a counterargument that directly weakens the force of the opponent's preceding argument. The proportion of Counter-C usage for the intervention group increased from 20.67% (SD = 29.67%) at initial assessment to 40.91% (SD = 19.28%) at final assessment, while the control group increased nominally from 18.84% (SD = 28.45%) at pretest to 19.93% (SD = 32.44%) at posttest. A significant group X time interaction was found, $F(1, 46) = 4.89$, $p=.032$, partial $\eta 2 = 0.096$.

### Intra-dyad Discourse Analysis

Analysis of the intra-dyad discussions is based on the 163 intra-dyad discussion transcripts of the discussion between members of a dyad as they participated in the sessions of the intervention. A coding scheme was developed to distinguish meta-level discursive moves about strategies (planning, evaluating, or predicting) from simply discussing the content of the arguments (see Table 1).

Table 1: Intra-dyad discourse coding scheme

| Code | Examples |
|---|---|
| **METACOGNITIVE STRATEGIES** | |
| PLANNING [PL] | -We could mention how effective the expulsion will be towards the kids in the school. Maybe we explain that better?<br>-I meant we can use the friends reason.<br>-What should we say to that? |
| PREDICTING [PR]: Predicting effect or consequence of strategy that hasn't happened | -They are still writing. What do you think they are going to talk about?<br>-They will say that he could learn English in school or they could get him a Spanish teacher.<br>-Don't know how they will come back from that |
| EVALUATING [EV]: Analysis of performance and strategy effectiveness | -We stuck a hole in their argument<br>-We gave them a good comeback<br>-They proving our point! Look what they keep writing!<br>-What's our main argument? |
| **METACOGNITIVE KNOWLEDGE** [MC] | -I don't even know why I'm arguing with this side. I got so many ideas in my head |

| | |
|---|---|
| | -Can we just talk to everybody instead of typing?<br>- You know what, we can use this! We can use this! This is our comeback! Our Counter-Counter argument. |
| **NON-METACOGNITIVE** | |
| NULL [N] | -Oh man |
| OFF TASK [OT] | -I'm MAD hungry. How much time til lunch? |
| TECH [TC] | -They just responded<br>-I don't think our network is connected |
| READING FROM SCREEN [RS] | -They said, "Aris needs to make friends" |
| TYPE-ALOUD [TA] | -We… think… the… boy… should… be… home… schooled |
| DICTATION [DT] | -…receiving a fatal warning, F-A-T-A-L. |
| DIVISION OF LABOR [DL] | -Let me type now |
| ARGUMENT DISCUSSON [AD]<br>Discussion of content (not discourse) | -He could hire a tutor |

### Relation between Meta-Level Intra-Dyad Discourse and Strategic Gain

To examine the relationship between meta-level discourse and strategic performance from initial to final individual assessment, students were categorized into two additional groups for each topic. The strategic improvement categorization is based on whether they increased or didn't increase with respect to a key indicator of strategic improvement, the proportion of Counter-C utterances from initial to final assessments.

For Topic 1, six students (three dyads) were dropped from the analysis of the first topic due to not participating in the intervention. Simply put, the partners didn't interact with one another during the entire first topic of the intervention. In consultation with the classroom teacher, these six participants were arranged in better pairs for the second and third topics. Of the 19 participants, 14 were classified as showing improvement from initial to final assessments. Of the 14 that improved, eight were also classified in the high-meta intra-dyad category, i.e., they were equal to or above the median for total number of meta-level utterances, and six were classified in the low-meta category. Of the five that didn't improve, two were above or equal to the median and three were below. While suggestive, these proportions were not significantly different ($a$ = .628, Fisher's Exact Test).

For Topic 2, of the 25 participants, 19 improved from initial to final assessments and six didn't. Of the 19 that improved, 13 were above or equal to the median for proportion of meta-level utterances and six were below. Of the six that didn't improve, all were below. These proportions were significantly different ($a$ = .005, Fisher's Exact Test).

For topic 3, of the 25 participants, 19 improved from initial to final assessments and six didn't. Of the 19 that improved, 13 were above or equal to the median for proportion of meta-level utterances and six were below. Of the six that didn't improve, all were below. These proportions were significantly different ($a$ = .005, Fisher's Exact Test).

### Relation between Total Utterances and Strategic Improvement

A similar analysis was conducted to determine if total intra-dyad talk related to strategic improvement. Two groups were formed per topic – the *high-talk* group, at or above the median, and the *low-talk* group below the median. The relationship between total talk frequency and strategic improvement, even while suggestive for topics 2 and 3, was not significant for any of the topics. Thus, it appears to be meta-level talk specifically, rather than talk in general, that is predictive of strategic gain.

## Discussion

The present results confirm that argumentation skill can be developed among young adolescents over the course of a goal-based, collaborative activity offering dense opportunities for engagement in discourse. Consistent with earlier findings (Kuhn et al., 1997, 2008; Kuhn & Udell, 2003), the intervention was effective in enhancing students' dialogic argument skills in addressing a new topic that had not been included in the intervention. Students in the intervention group exhibited an increased proportion of usage of skilled argument strategies (Counter-C) and a decreased frequency of usage of less advanced (Exposition) strategies, compared to the control group.

These findings are inconsistent with claims that argument skills are entirely domain specific (Stein & Miller, 1993). Participants in the present study only engaged each other on the capital punishment (CP) topic twice – once before the intervention as a pretest and once at the end as a posttest. The pretest did provide students with some experience with the topic, but no more than it did students in the control group. Improvements within a domain could be due to acquisition of specific content-bound arguments within the domain, but transfer to a different domain requires a more complex account.

It warrants emphasis that this progress was observed in a group of significantly academically disadvantaged students. Prior to the intervention, these young adolescents had little occasion to practice sustained argumentive discourse in their daily lives. Observation of their classroom experience showed that they experienced only infrequent opportunities to express their views and then only in response to a teacher's question, rather than in debate with one another. Yet these students readily engaged the activity and the majority showed clear advance in skill.

The contribution of the present study lies in the insight it offers regarding the mechanisms that make the method effective. The intervention created conditions in which focused peer discourse was required and supported. The results indicate that its value lay not in talk per se, the quantity of which did not predict improvement. Instead, it has been proposed here, a major value of such discourse is its support of meta-level representation and regulation of discourse processes. The findings support this hypothesis. Students producing a high proportion of meta-level utterances during peer discourse tended to make the greatest progress in individual argument skill. This relationship was statistically significant for the latter two of the three phases of the intervention period. During the first topic, we can speculate, students were settling into an activity that was very new to them and gains were less consistent overall.

The present results, it should be emphasized, don't simply show that students displaying a high proportion of meta-level utterances did better on the argument skill posttest. Rather, frequent expression of meta-level utterances with a collaborating partner is associated with advance in individual argument skill. This advance, note further, could be observed across the range of initial skill levels. Initially less competent and initially more competent individuals both improved when they engaged in frequent meta-level talk.

Why might engaging in meta-level regulatory discourse lead to skill advance? Meta-level understanding of the goals of argumentation and of strategies to achieve these goals develops over time with engagement and practice. Such meta-level discussions aren't simply reflecting the individual's current understanding of task and strategies. The discussion is part of the process, along with engaging the opposition, of constructing the understanding.

The most useful and likely accurate way to conceptualize the relationship between meta-level understanding and strategic engagement with a cognitive task is most likely as one of mutual bootstrapping (Kuhn & Pearsall, 1998). Implementation and practice of strategies leads to greater understanding and valuing of them, in bottom-up fashion, while top-down monitoring and management of strategies plays an increasing role. This is not a new idea, but the present study contributes to our understanding of how this process may work by examining and identifying differences between cases in which it works well and less well over an extended period of time. The Counter-C strategy was used in combination with less advanced strategies over the course of the intervention. Both the effective use of advanced strategies and ineffective use of less-advanced strategies may "feed-up" to the meta-level and strengthen awareness of the utility of each strategy. Meta-level understanding also "feeds down" to the strategic level. Increasing awareness of the coordination of task objectives and effective strategies to reach task objectives guides strategy implementation.

As microgenetic research has repeatedly shown, enhanced meta-level awareness of strategies, and particularly more advanced strategies, doesn't mean that less-advanced strategies cease to be used (Siegler, 2006). More likely, however, meta-level understanding was under construction throughout the intervention through repeated engagement rather than emerging in a single observable moment of change.

In summary, the present results support the conclusion that the opportunity for explicit meta-level discourse contributes to meta-level understanding, which has been shown here to relate to the development of strategic skills. We are far from a complete understanding of the mechanisms involved and more observation of them in action is needed. In particular, the need to understand how top-down and bottom-up processes reciprocally support one another will arguably require more attention to individuals' understanding of their practices, rather than simply performance of them.

## References

Anderson, R., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., et al. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction*, 19(1), 1-46.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116).

Hillsdale, NJ: Lawrence Erlbaum Associates.
Chinn, C. A. (2006). Learning to argue. In A. O'Donnell, C. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology*. Mahwah, NJ: Erlbaum.
DeFuccio, M., Kuhn, D., Udell, W., & Callender, K. (2009). Developing argument skills in severely disadvantaged adolescent males in a residential correctional setting. *Applied Developmental Science, 13*(1), 30-41.
Felton, M. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development, 19*, 35-52.
Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skills. *Discourse Processes, 32*, 135-153.
Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*, 906–911.
Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven: Yale University Press.
Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science, 6*, 133-139.
Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178-181.
Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the Computer: A Microgenetic Study of Developing Argument Skills in a Computer-Supported Environment. *Child Development, 79*(5), 1311-1329.
Kuhn, D., & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance. *Cognitive Development, 13*, 227-247.
Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction, 15*(3), 287-315.
Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*(5), 1245-1260.
Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking and Reasoning, 13*, 90-104.
Nussbaum, E.M. (2003). Appropriate appropriation: Functionality of student arguments and support requests during small-group classroom discussions. *Journal of Literacy Research*, 34, 501-544.
Nussbaum, E.M., & Kardash, C. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, 97, 157-169.
Nussbaum, E.M., & Sinatra, G. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology*, 28, 384-395.
Reznitskaya, A., Anderson, R., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes*, 32, 155-175.
Sampson, V. & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, 93(3), 448-484.
Siegler, R. (2006). Microgenetic studies of learning. In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2.* Cognition, perception, and language. (6th ed.). Hoboken, NJ: Wiley.
Stein, N., & Miller, C. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glasser (Ed.), *Advances in instructional psychology*. Hillsdale, NJ: Erlbaum.

# Eliciting and Developing Students' Ideas and Questions in a Learner-Centered Environmental Biology Unit

Christopher J. Harris, SRI International, 333 Ravenswood Avenue, Menlo Park, CA, 94025
christopher.harris@sri.com
Rachel S. Phillips, University of Washington, 312 Miller, Box 353600, Seattle, WA, 98195
rachelsp@u.washington.edu
William R. Penuel, SRI International, 333 Ravenswood Avenue, Menlo Park, CA, 94025
william.penuel@sri.com

**Abstract:** An important instructional practice for teachers in learner-centered science classrooms is to be able to work productively with students' ideas and questions. Eliciting and thoughtfully attending to students' thinking in order to help them advance in their learning is no easy matter, however. This study examined how teachers enacted instruction to elicit, connect, and build upon their elementary students' ideas and questions during an innovative twelve-week learner-centered environmental biology unit. The purpose of the study was to identify and describe successes, issues, and challenges related to incorporating students' ideas and questions into science instruction. Primary data sources included field notes taken during classroom observations and teacher interviews. We present three contrasting cases of teachers to highlight evidence that shows teachers' differing strategies for eliciting students' ideas and questions, and for developing students' ideas, questions and questioning skills. We discuss practical implications for designers of inquiry-based science curricula and professional development.

## Introduction

Learning sciences research has identified as a key design principle that learning environments must be "learner-centered," that is, they must attend to and make use of what students bring to the classroom learning situation (National Research Council [NRC], 2000a). By eliciting and attending to what students bring, teachers can actively engage students' ideas about and orientations to what and how they are learning (Schwartz & Bransford, 1998). Moreover, making use of students' contributions in teaching can help transform students' discipline-oriented thinking and ways of participating in disciplinary practices (Baranes, Perry, & Stigler, 1989).

To support student sense-making effectively in learner-centered science environments, teachers need to shift from the typical pattern of talk in classrooms (e.g., Cazden, 1988; Mehan, 1979) to new forms of discourse that promote productive thinking and participation. In addition, teachers need to be able to make creative use of the diversity of student ideas during instruction. This entails making productive adaptations to students and their contributions as instruction unfolds. A challenge for designers is to develop flexibly adaptive curriculum materials (Schwartz, Lin, Brophy, & Bransford, 1999) that provide the structure for teachers to enact ambitious science instruction while at the same time provide support and room for teachers to make use of what students bring to the learning context. Even with such materials, teachers are likely to diverge in their approaches to and success in creating richer forms of discourse in classrooms.

In this study, we examined teachers' instructional moves to elicit and develop students' ideas and questions as they enacted a twelve-week learner-centered environmental biology unit with their fifth grade students. The unit was designed specifically to help teachers reconfigure their classrooms for science learning in ways that were more learner-centered, with a particular focus on creating opportunities for teachers to elicit and work with students' ideas and questions as a source for student-led investigations of habitats. A central aim of our study was to gain insight into the ways in which teachers' enactments enabled them to work with students' ideas and questions to help advance learning. To this end, we present three contrasting cases of teachers to highlight evidence that shows teachers' differing strategies for eliciting students' ideas and questions, and for developing ideas, questions and questioning skills.

## Theoretical Framework

Formulating and refining questions that can be answered through scientific investigations is a fundamental ability required for science inquiry (NRC, 1996, 2000b, 2007). Developing scientific questions is a creative act at the heart of scientific activity (Shodell, 1995). In practice, however, student questions are rare in classrooms, compared with teacher questions (Carlsen, 1993; Dillon, 1988). Moreover, students find it challenging to take an ill-formulated question posed at the outset of an inquiry and clarify or refine it in ways that can productively guide inquiry (Krajcik, Blumenfled, Marx, & Soloway, 2000; Lehrer, Giles, & Schauble, 2002; van Zee, Iwasyk, Kurose, Simpson, & Wild, 2001). Thus, a critical challenge is to help teachers use curriculum to enable students to both to generate and develop questions that can guide inquiry.

Research on student question generation suggests that teachers' actions are critical. When teachers establish discourse patterns (e.g., brainstorming, "K-W-L" sequences) that elicit student questions and engage students in discussion about their observations of familiar contexts, more questions are likely to emerge in the science classroom (Penuel, Yarnall, Koch, & Roschelle, 2004; van Zee, et al., 2001). In these discussions, teachers' use of "wait time" and reticence from immediately judging student contributions give space for more student questions to emerge (Gallas, 1995; Rowe, 1986). Besides observation, other sources of student questions that teachers can tap are prior knowledge, cultural beliefs, information from media, and family experiences (Chin & Chia, 2006; King, 1994; Reeve & Bell, 2009).

Not all student-generated questions can be investigated in classrooms and those that can be pursued often require the assistance of the teacher and peers to develop. Some questions students pose can be researched using books or digital resources and are answered readily with reference to facts; others are "wonderment" questions that invite hypothesis generation and prediction (Scardamalia & Bereiter, 1992). Teachers can encourage the latter type of questions by engaging students in extended problem solving activities (Chin, Brown, & Bruce, 2002) and in student-led investigations (Hofstein, Navon, Kipnis, & Mamlok-Naaman, 2005). In addition, putting students into collaborative learning situations where they must clarify questions and design investigations together can help develop students' questions (Marbach-Ad & Sokolove, 2000). Finally, teacher discourse moves such as "revoicing" that explicitly align different student contributions and ideas to content and to set up contrasts among students helps develop questions provides a motivating context for the development of questions (O'Connor & Michaels, 1993).

## Curricular Context

The research described here is part of a larger design-based research project that included the creation of an elementary school science inquiry unit as a means to support science instruction that is challenge-based (Schwartz, Lin, Brophy, & Bransford, 1999), leaner-centered (NRC, 2000a), and authentic in the sense that students engage in tasks that are relevant to the science topic under study and to their own lives and interests. This curricular unit was created via a school-university partnership, subsequently piloted by teachers in the district, and then revised collaboratively with teachers and researchers. Cornerstones of the unit were student-choice and student-driven inquiry related to an over-arching challenge, all couched in socially interactive group work. The newly revised unit was then systematically studied as it was being implemented by a cadre of teachers in their elementary classrooms, which exemplifies the tenets of a research into practice, practice into research approach (Design-Based Research Collective, 2003). In this curriculum unit, titled *The Isopod Habitat Challenge* (IHC), students move through phases of inquiry as a means to solve their ultimate challenge—to create an optimum habitat for isopods. Students first share their ideas and questions about isopods, take part in a teacher-guided investigation, and then participate in student-generated investigations to answer their questions. As students move through each phase, they revisit and revise their initial ideas and questions, conduct new research, revise their ideas and reformulate their questions again, and then present their final habitat plan in a public forum.

The unit materials included features that are *educative* (Davis & Krajcik, 2005), meaning features meant to help teachers learn, to better support student learning. Educative features in the materials included overviews of the instructional stance, organization, and phases of the unit, flexible pacing guides that supported teacher planning, and embedded notes that signified decision points (i.e., critical junctures) during instruction where it was especially important to attend to student thinking. The lesson materials also included prompts such as idea generating (e.g., brainstorming) and questioning prompts that were meant to help teachers elicit student ideas and questions. Materials for students included an *idea journal* for developing and recording ideas and questions and for iteratively investigating key questions as they progressed through phases of inquiry, as well as a *team planner* for collaboratively designing the habitat for their isopods. Teachers had access to the unit materials online through the school district website.

University researchers and two teachers who initially piloted the unit led three formal professional development sessions to support teachers' learning of how to use the curriculum materials with students. All of these sessions took place after school hours in participating teachers classrooms. The focus of the professional development sessions included instruction on science content, descriptions of lessons and activities, modeling of lessons and activities through video clips, pacing of the unit, classroom management, and discussion on effective instructional strategies for teaching science as inquiry. Time was also spent troubleshooting challenges that teachers encountered during unit enactment as well as celebrating and sharing successes. In addition to the formal professional development sessions, the teachers were in close contact via email and in-person meetings at school sites with university researchers who were involved in the unit development. This frequent electronic and in-person communication, or "just in time" professional development provided individualized support to the teachers while they were enacting the unit in their classrooms.

## Methods

The overarching research question guiding this study was, *For a curriculum unit that aims to develop students' skill in posing scientific questions, how do teachers vary in the ways they elicit, re-voice, connect, and/or build upon students' science ideas and questions?* We approached the question by focusing, as past research has, on teachers' instructional moves to elicit and help students to go further with their ideas and questions. By analyzing differences among teachers using the same curriculum materials, our comparative case approach enabled us to examine how teachers' enactments shaped their students' learning opportunities.

### Setting and Participants

A total of eight teachers across five elementary schools in one mid-size suburban school district in the Pacific Northwest enacted IHC over 12 weeks in their fifth grade classrooms. We present case studies of three of the teachers from three different schools in the district. The three teachers, Ms. Atwell, Mr. Jimenez, and Ms. Lesh (pseudonyms), were purposefully selected as contrasting cases for the present study. Mr. Jimenez was a M.A.-level teacher in his $15^{th}$ year and taught at an elementary school whose student population was diverse (24% Asian American, 40% Caucasian, 22% Hispanic, 6% African American, and 8% multi-ethnic), with a high percentage of students receiving free or reduced price lunch (47%). Ms. Atwell was a M.A.-level fourth-year teacher who taught at an elementary school that was predominately Caucasion (60%) and Asian American (19%), with a small percentage of students receiving free or reduced price lunch (13%). Ms. Lesh was a B.A.-level fourth-year teacher at an elementary school that was comprised primarily of Asian American (52%), Hispanic (19%) and Caucasian (17%) students, with 29% of students receiving free or reduced price lunch.

Within the sample, the three teachers' students represented three different levels of accomplishment with respect to the unit's goals. Students in Ms. Lesh's class had the highest average score on the post-assessment and Mr. Jimenez' students scored lowest. Ms. Atwell's students scored in the middle. Preliminary analyses of observation data suggested these three teachers also differed with respect to their approaches to implementing the unit, such that the differences in student results might be analyzed in terms of these differences.

### Procedures

Data sources included narrative documents of lessons produced by integrating field notes and observation protocols completed by classroom observers, as well as semi-structured teacher interviews conducted by researchers during and after teachers' enactment of the unit. The classroom observations were spread across the 12-week unit, enabling observers to visit classrooms and record lessons at the beginning, middle, and end of the unit. We used as the basis for analysis a set of narrative documents from 18 classroom observations across the three teachers. The narratives were comprehensive descriptions of classroom events, targeting teacher and student actions, interactions, and conversations.

In analyzing narratives, we developed and employed a coding scheme based on discourse interactions, focused particularly on patterns of teacher "uptake" (Nystrand & Gamoran, 1991) of student ideas and questions during instruction. The coding scheme was developed through an iterative process of creating codes based on hypotheses, coding evidence of elicitation and uptake, modifying and refining codes, and recoding consistent with recommendations for qualitative data analysis by Miles and Huberman (1994). Independent coding of narratives was conducted with two coders who met regularly to compare evidence for codes and calibrate their approaches for identifying evidence. Differences were resolved through discussion and consensus. In instances where both coders were uncertain about a piece of evidence, a third researcher provided judgment to help clarify and reach agreement. A short statement of justification was written for each piece of evidence that linked the evidence to a code.

Interviews were conducted with each teacher at two time points: once midway through the unit and a second time at the conclusion of the unit. Specific to the present study, the semi-structured interviews addressed teachers' perspectives on students' experiences, ideas, and questions during instruction, and included questions on what teachers did in response to student contributions. For example, teachers were asked whether and how students' ideas influenced their teaching, the kind of guidance they provided to students during student-led research activities, and how they tried to respond to students' problematic ideas. Interviews were audio recorded, transcribed, and then used to further explore developing patterns and themes that emerged from the analyses of narrative data.

### Constructing the Cases

We sought to select cases that, through analysis, could provide guidance to teachers about how to work productively with students' ideas in ways that went beyond simple elicitation. To this end, we took an explanatory, multiple-case study approach (Yin, 2003) that focused on explaining variation in individual teacher enactment from curriculum use. In this approach, researchers take as their aim not simply to describe the phenomena under study, but to seek out explanations for why cases unfold the way they do. Yin's (2003) recommendation is that researchers develop a set of initial possible "rival explanations" for patterns in the data that they expect to find, both as a guide to instrument design and as a method for guarding against confirmation bias.

## Results

In this section we highlight evidence from our analyses focused on teachers' strategies for eliciting students' ideas and questions, as well as strategies for developing ideas, questions and questioning skills. We frame these strategies as *instructional moves* – actions meant to facilitate learning typically through a combination of speech and gesture. *Strategies for eliciting* refer to instructional moves made by teachers in an effort to draw out and make student ideas and questions visible. *Strategies for developing* pertain to the "next step" instructional moves made by teachers to respond to student thinking. Though we identify and describe instructional moves separately, we recognize that they are oftentimes intertwined during instruction and that teachers typically enact sequences of moves to support learning.

Overall, our findings show that all three of the teachers were effective in employing instructional moves to elicit student ideas and questions. Moreover, their strategies for eliciting were similar. However, teachers differed widely in their next step instructional moves and how they worked with student ideas and questions to help students go deeper in their thinking; the teacher whose students scored the lowest on the student assessments made the fewest of these kinds of moves. To illustrate these similarities and differences, we present below summaries of teachers' moves with examples that characterize their instruction.

### Strategies for Eliciting Ideas and Questions

All three teachers used the elicitation prompts in the unit materials as well as their own strategies. Among the three teachers, we identified two common strategies for elicitation – posing questions to elicit students' ideas and questions, and inviting students' ideas and questions. The teachers used these strategies to elicit both procedural and wonderment thinking. That is, teachers asked and invited procedural questions that addressed basic information about how to accomplish teacher-structured tasks as well as how to carry out student-generated ones. They also asked and invited wonderment questions that dealt with hypothesizing and predicting, explaining and clarifying, and making sense of investigative experiences and results.

#### Mr. Jimenez

Mr. Jimenez tended to pose questions to elicit student thinking about procedural knowledge for engaging in tasks, such as what materials to use for investigations, the steps to follow, and how to record work. During activities and investigations, Mr. Jimenez asked basic information questions to check in on students' progress (e.g., what are you doing?) and help students in carrying out their tasks (e.g., what will you do next?) When students experienced difficulties in carrying out their procedures, he tended to pose elicitation questions to help students move forward in their investigative process (e.g, how does that help answer your question?).

Mr. Jimenez also engaged in a regular elicitation pattern of inviting students to share their group work experiences and outcomes. He tended to structure these opportunities as whole-class reporting sessions in which students from different investigative groups take turns stating their group ideas, questions, and findings. Elicitation invitations included such prompts as "what did you research?", what did you find out?" and "let's have each group tell us what they did". In the following excerpt, Mr. Jimenez invited students to report their findings as he recorded them on a smart board:

S1: They live under rocks and in soil.
S2: When you touch them they roll up...
T: Okay… what was the other comment you said?
S: They roll up.
T: Oh, okay, they roll up.
S3: They are in different segments.
S4: They have eyes.
S5: They don't have eyes!!
T: There is some disagreement, let's write them both down for now. If you have any other ideas, please share.

## Ms. Lesh

Ms. Lesh posed procedural and planning questions to her students to ensure that they clearly understood how to do tasks and to help them in the planning process of their investigations. During whole class discussions about planning procedures for investigations, Ms. Lesh tended to invite students to help determine steps and reason through the benefits or drawbacks of following particular steps. For example, while collectively planning an experiment in which they were going to examine isopods' food preferences, Ms. Lesh elicited students' ideas regarding the number of isopods needed for the experiment by asking, *how many isopods do we need to be able to answer our question?* Students then offered their ideas and reasons for various numbers. One student's response to the elicitation was to select either 3 or 5 isopods, *"because if isopods moved in even numbers to both sides, it wouldn't be possible to decide."*

When students encountered difficulties during small-group activities and investigations, Ms. Lesh typically elicited their ideas about how the problem/issue might be resolved, rather than offering an immediate solution. She also monitored groups closely, encouraging anticipatory thinking by eliciting students' predictions about what might happen next as they were proceeding through investigations. She tended to ask students while they were working in groups to explain what they were doing.

After group investigations, Ms. Lesh tended to have students report out their results as well as their successes and challenges in carrying out their investigations. Of note is that Ms. Lesh elicited wonderment questions primarily when prompted to do so in the materials (e.g., prompts in the materials to generate student-driven questions about isopods).

## Ms. Atwell

Ms. Atwell posed questions to elicit students' procedural knowledge for carrying out tasks as well as their ideas for planning and conducting investigations. Similar to Ms. Lesh, she tended to ask students to report out on their designs for student-led investigations and then elicit from students their reasoning for their design decisions (e.g., why did you decide to do it like that?). She also tended to pose questions to elicit students' ideas about conceptual topics. These elicitation questions were typically framed as wonderment questions (e.g., what do you think?) that probed students' grasp of science ideas (e.g., what is a habitat – what are your ideas?), their ideas about what makes for a good researchable question (e.g., what does a scientific question involve?), and the kinds of questions that students were working with (e.g., what is your question?).

She also elicited conceptual questions from students and structured discussions in a manner that encouraged students to clarify (e.g., what do you think she means?), build upon (who can add to what was said?), or counter one another's ideas (e.g., who has a different idea about how we can set this up?). After investigations, her whole class elicitations tended to focus on what students learned (e.g., what did you find out?) as well as what new questions they had (e.g., what questions do you have now?).

## Strategies for Developing Ideas, Questions and Questioning Skills

All three teachers worked with their students to generate, refine and pursue questions that served to frame students' investigations and guide their inquiries. In addition, students in all three classes appeared to benefit when teachers made suggestions and pressed students to clarify and refine their questions and investigations. However, the three teachers differed in their level of attention to students' ideas, questions, and questioning skills. Here we highlight the diverse ways in which teachers attended to student thinking.

## Mr. Jimenez

Mr. Jimenez tended to incorporate student thinking into whole-class activities and discussions, but often did not go beyond simple acknowledgement of students' contributions. There were very few instances where he tried to really work with students on developing their ideas. When he did try to do so, it tended to be in the context of supporting students in designing investigations. In the following excerpt, Mr. Jimenez tried through his elaborations and questions to support students in thinking about controlling variables in their procedures:

> T: So, what materials do you think are necessary to do this investigation or experiment?
> S: Soil.
> T: Soil, okay (writes on "soil" on board). About how much soil do you think is necessary for this?
> S: A handful.
> T: Okay, a handful, uh, so, if we just say a handful and we just put it in the runway, how are we going to know if the other side is going to have the same amount, okay? So, what can we do with the soil?
> S2: Use cups.
> T: Use cups, okay? Okay, so that's important, so we need to keep the amount of soil the same, that's another variable that we need to keep the same.

Another way that Mr. Jimenez incorporated student thinking into whole-class activities was by having groups report out on their group activities and learning, and asking individual students to share ideas and questions. But when students shared results, posed a question, or raised an issue, he tended to simply acknowledge the contribution and would then move on to another topic. For example, when Mr. Jimenez encouraged groups to bring out their ideas, he often responded to contributions by saying “next group”, “okay”, and “so, now we are going to…”. In this way, Mr. Jimenez structured discourse so that ideas and questions were elicited but fell short on leveraging those contributions to help deepen students’ thinking.

## Ms. Lesh

Ms. Lesh actively worked with students to ensure that their questions and procedures were feasible. She regularly pressed students to refine their questions, often engaging individual students in elaborating on their questions so she could then help tighten them. When pressing students to be detailed in their procedures, she explained that as scientists “you have to be exact” and regularly reminded them to “put in more detail” and “be more specific”. When supporting students in refining their research questions, she tended to revoice their questions and press them to elaborate. Through her questions of their research questions, students were prompted to clarify and make their questions more feasible for classroom investigation.

During end-of-investigation discussions, Ms. Lesh connected student findings and insights to the work of other groups when the opportunity arose. For example, when a student reported observations that were similar to the work of another group, Ms. Lesh made explicit the similarities in findings. When Ms. Lesh took up students’ conceptual ideas, she tended to value and validate their ideas with supportive comments like “you’ve got it” and “excellent” and expressing excitement through gestures such as high fives. When students were off the mark in their ideas, she tended to try and redirect student thinking through clarification questions (e.g., are you really asking about day or are you interested in light vs. dark?) and explaining her interpretation of students’ ideas.

## Ms. Atwell

Ms. Atwell worked continuously with students to help them learn how to ask researchable questions. In the following excerpt, she asked questions to help a student refine a question, and also used the conversation to model for the class how she was thinking when listening to what students were saying:

> S: what does the climate need to be for them?
> T: what do you mean by climate?
> …
> T: (to whole class) What I’m trying to do is paraphrase in my mind what I think she’s saying. It’s a good thing for you to do when you’re listening, too.
> T: (to student who asked question) We have temperature questions, but it seems like you’re going a little further.
> S: Does the isopod’s habitat need to be hot and dry or cold and moist?

Ms. Atwell held inquiry discussions and structured them to involve students in clarifying their own ideas and questions and exploring others’ perspectives. She also invited students to help each other think through their ideas and questions. For example, when a student shared her prediction about isopods preferences for location, Ms. Atwell asked the class, “does anyone want to respond?” She then allowed students to ask clarification questions regarding the initial prediction and encouraged students to provide reasons why they did or did not think the prediction would hold true. In this way, she established a discourse environment in which students saw their task as sense making and knowledge building.

Similar to Mr. Jimenez and Ms. Lesh, Ms. Atwell asked groups to report out their designs for their student-led investigations and, once investigations were conducted, encouraged groups to report their findings. Differently, however, Ms. Atwell tended to pose questions, encourage students to pose questions, and gave clear feedback when recognizing student thinking. For example, when a student group was describing the reasoning for their second research design, both Ms. Atwell and a student engaged the group:

> S1 (G1): Since in the first investigation they liked the moist side…
> T: Good thinking – I like how you built on knowledge you already had.
> S2: I have a question, why did you decide to do it like that?
> S1 (G1): We wanted to have it super light and super dark on each side.
> T: So, were you thinking if you had more contrast you might get better data?
> G1: Yes *(multiple students respond).*
> T: Thanks team.

## Discussion and Conclusions

Our comparative case studies revealed that the three teachers differed in the ways they developed or took up student questions posed in investigations in ways that potentially help explain observed differences in student learning across classrooms. In Mr. Jimenez' class, where students scored lowest on a test of their skill in posing investigable questions, the teacher made few moves to help develop student thinking. By contrast, in the two other classrooms where students did better on the assessment, the teachers used multiple strategies for developing student questions. Even when implementing the same curriculum unit, which provided explicit guidance about how to elicit questions and called for multiple iterations in students' design of investigations, there was significant variation in how teachers took up the materials.

Implementation variation of this kind is hardly surprising (see, for example, Schneider, Krajcik & Blumenfeld, 2005), but the particular ways teachers varied and were similar in this study points to specific aspects of developing students' questioning skills that future curriculum materials may need to consider. Strikingly, all three teachers (as well as others we observed in this study) were able to use the curricular guidance in the materials to elicit student questions. The teachers in interview data, not reported here, all found this aspect easy and interesting to enact. By contrast, teachers had a harder time consistently helping students develop their ideas. Sometimes, as we report elsewhere (Phillips, Harris, Penuel, & Cheng, 2010), it was because teachers had to balance considerations of time with the need to give students multiple opportunities to develop their thinking. But in the successful teachers' classrooms, we also found evidence of classroom norms developed through sequences of moves described above. The individual moves employed were in some cases simple and straightforward, but orchestrating them in class was a skill that varied from teacher to teacher in the study. A single move can be described and pointed out easily, but learning how to orchestrate them into coherent, effective sequences of "next step" moves takes time and is likely to develop primarily through reflective practice.

One conclusion from our findings is that the IHC curriculum materials must provide greater educative support for teacher learning to develop students' questions. The team that designed the unit shared our conclusion and developed additional resources to support future teachers' enactment that illustrate ways to develop teacher questions. Some of these resources included video exemplars from the two teachers in the study who were more successful in developing student questions. For its part, the district in the study plans to take the results of our implementation analysis and share our list of discursive moves with science instructional coaches. Their plan is to use the moves as a lens for observing and mentoring teachers to make the most of the curriculum materials.

Future research is needed on strategies to develop students' questioning skills that compare the efficacy of different sequences of moves for developing questions. Our analysis did not permit a fine-grained analysis to distinguish how successful the two teachers who had a broader repertoire of moves for developing questions were. It may be that some strategies are more effective with some students, depending, for example, on their prior experience with planning investigations. Given that questioning is such a fundamental aspect of science and most students remain novices at posing and investigating their own questions, research in this area is an important focus for learning scientists to pursue.

## References

Baranes, R., Perry, M., & Stigler, J. W. (1989). Activation of real-world knowledge in the solution of word problems. *Cognition and Instruction, 6*(4), 287-318.

Carlsen, W. S. (1993). Teacher knowledge and discourse control: Quantitative evidence from novice biology teachers' classroom. *Journal of Research in Science Teaching*, 30(5), 417-481.

Chin, C., Brown, D. E., & Bruce, B. C. (2002). Student-generated questions: A meaningful aspect of learning in science. *International Journal of Science Education*, 24(5), 521-549.

Chin, C., & Chia, L.-G. (2006). Problem-based learning: Using ill-structured problems in biology project work. *Science Education*, 90(1), 44-67.

Davis, E., & Krajcik, J. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3-14.

Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Dillon, J. (1988). The remedial status of student questioning. *Journal of Curriculum Studies*, 20, 197-210.

Gallas, K. (1995). *Talking their way into science: Hearing children's questions and theories, responding with curricula*. New York: Teachers College Press.

Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, 42(7), 791-806.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31(2), 338-368.

Krajcik, J. S., Blumenfled, P. C., Marx, R. W., & Soloway, E. (2000). Instructional, curricular, and technological supports for inquiry in science classrooms. In J. Minstrell & E. H. v. Zee (Eds.), *Inquiring into Inquiry Learning and Teaching in Science* (pp. 283-315). Washington, DC: American Association for the Advancement of Science.
Lehrer, R., Giles, N. D., & Schauble, L. (2002). Children's work with data. In R. Lehrer & L. Schauble (Eds.), *Investigating real data in the classroom: Expanding children's understanding of math and science* (pp. 1-26). New York: Teachers College Press.
Marbach-Ad, G., & Sokolove, P. G. (2000). Can undergraduate biology students learn to ask higher level questions? *Journal of Research in Science Teaching*, 37, 854-870.
Mehan, H. (1979). *Learning lessons: Social organization in the classroom.* Cambridge, MA: Harvard University Press.
Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.
National Research Council (1996). National Science Education Standards. Washington, DC: National Academy Press.
National Research Council (2000a). *Inquiry and the National Science Education Standards*. Washington, DC: National Academy Press.
National Research Council (2000b). *How people learn: Brain, mind, experience.* (Expanded ed.). Washington, DC: National Academy Press.
National Research Council (2007). *Taking science to school: Learning and teaching science in Grades K-8*. Washington, DC: National Academy Press.
Nystrand, M., Wu, L. L., & Gamoran, A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes, 32*(5), 135-198.
O'Connor, M. C., & Michaels, S. (1993). Aligning academic talk and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly*, 24, 318-355.
Penuel, W. R., Yarnall, L., Koch, M., & Roschelle, J. (2004). Meeting teachers in the middle: Designing handheld computer-supported activities to improve student questioning. In Y. B. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon & F. Herrera (Eds.), *Proceedings of the International Conference of the Learning Sciences* (pp. 404-411). Mahwah, NJ: Lawrence Erlbaum.
Phillips, R. S., Harris, C. J., Penuel, W. R., & Cheng, B. (2010). *Teachers managing students' ideas, questions, and contributions in the context of an innovative inquiry-based elementary science unit.* Paper to be presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, PA.
Reeve, S., & Bell, P. (2009). Children's self-documentation and understanding of the concepts 'healthy' and 'unhealthy'. *International Journal of Science Education*, 31(14), 1953-1974.
Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, 37(1), 43-50.
Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. Cognition and Instruction, 9, 177-199.
Schneider, R. M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of teacher enactments in reform classrooms. *Journal of Research in Science Teaching, 42*(3), 283-312.
Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475-522.
Schwartz, D. L., Lin, X., Brophy, S., & Bransford, J. D. (1999). Toward the development of flexibly adaptive instructional designs. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instruction theory* (Vol. 2, pp. 183-213). Mahwah, NJ: Erlbaum.
Shodell, M. (1995). The question-driven classroom: Student questions as course curriculum on biology. *The American Biology Teacher*, 57, 278-281.
van Zee, E. H., Iwasyk, M., Kurose, A., Simpson, D., & Wild, J. (2001). Student and teacher questioning during conversations about science. *Journal of Research in Science Teaching*, 38(2), 159-190.
Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

## Acknowledgments
This work was supported by grants from the Bill and Melinda Gates Foundation and the National Science Foundation (NSF #0354453). The opinions, findings and conclusions expressed here are those of the authors and do not necessarily reflect the views of the funders or the authors' institutions. We gratefully acknowledge Kari Shutt, Allison Moore, Kathryn Torres, Kersti Tyson, and Katie Van Horn of the University of Washington for their assistance with data collection and to Aisha Heredia and Sheila Shea of SRI International for their assistance with data analysis. We extend special thanks to the teachers and students who participated in the project.

# Mentor Modeling: The internalization of modeled professional thinking in an epistemic game

**Abstract:** Players of *epistemic games*--computer games that simulate professional practica—have been shown to develop epistemic frames: a profession's particular way of seeing and solving problems. This study examines the interactions between players and mentors in one epistemic game, Urban Science. Using a new method called *epistemic network analysis*, we explore how players develop epistemic frames through playing the game. Our results show that players imitate and internalize the professional way of thinking that the mentors model, suggesting that mentors can effectively model epistemic frames, and that epistemic network analysis is a useful way to chart the development of learning through mentoring relationships.

## Introduction

There is a broad consensus that mentoring is an important part of young people's intellectual, social, and emotional development (Freedman, 1999). In this paper, we look at one powerful form of mentoring—the mentoring that takes place in professional training—and examine how one feature of professional mentoring—modeling— can be used to help players in a computer game learn real-world skills, knowledge, values, and ways of thinking. The context for this study is the epistemic game Urban Science, a game modeled on a practicum course for graduate urban planning students (Shaffer, 2006b). Epistemic games are computer-based role-playing games that simulate professional training.

A critical element of professional practica is the interaction between young professionals and their mentors (Schön, 1983, 1987). In the urban planning practicum, students meet regularly with planning consultants: experienced professionals with whom the students talk about their work. These reflective conversations play a critical role in developing a professional way of thinking. In Urban Science, mentors--who are also called planning consultants--do the same for the players. As in the planning practicum, planning consultants in Urban Science help novices with the intricacies of the work, provide feedback, and inspire them when they are stuck.

A central role of professional mentors, however, is to model the way of thinking that is unique to their profession: what Shaffer (2006a) calls the epistemic frame. This paper looks at whether this same form of mentoring takes place in one epistemic game, and if so, whether it has the same effect as mentoring in the professional context. We ask to what extent players in an epistemic game imitate the epistemic frame that in-game mentors model, and whether this imitation leads to the internalization of a professional way of thinking. While the results are specific to this study, the type of mentoring and learning that happens in epistemic games could be useful models for designing new types of educational settings, in schools or elsewhere.

## Theory

There is a wide body of research on mentoring and its effects on youth development that focuses on the role that mentors can play in promoting positive behavior and habits in youth (DuBois & Silverthorn, 2005). In terms of cognitive development, Vygotsky (1978) is important for his work on how the development of the child's higher cognitive processes depends on the presence of mediating agents in the child's social learning environment. While Vygotsky himself "emphasized symbolic tools-mediators appropriated by children in the context of particular sociocultural activities, the most important of which he considered formal schooling" (Kozulin, 2003), scholars have extended the discussion of those mediating agents to emphasize mediation through another human being. Much of the literature on social mediation is concerned with how adults can facilitate the child's problem solving process by arranging and structuring the problem for them, and by participating in the problem-solving itself. Joint problem solving (Wertsch, 1978) and other situations where a child's current capabilities are extended through the support of an adult, are often framed as "scaffolded" learning (Wood, 1999) or as cognitive apprenticeship (Rogoff, 1990). Mediating strategies like scaffolding, while perhaps too context-dependent and numerous to be simply classified (Kozulin, 2003), have been well described (Schaffer, 1996), but how particular aspects of social mediation contribute to a child's cognitive development is unclear. How social learning processes are converted into internal developmental processes Vygotsky leaves murky, suggesting that showing "how external knowledge and abilities in children become internalized" is an important agenda.

In particular, he describes how children imitate problem-solving techniques. As Vygotsky (1978) puts it, "using imitation, children are capable of doing much more in collective activity or under the guidance of adults." Children can imitate adults or more advanced peers to handle problems that would otherwise be beyond them. As Valsiner and Van der Veer (1999) point out, Vygotsky's use of the concept of imitation is more sophisticated than simple copying; it is part of a learning and developmental process. When Vygotsky argues that children "can imitate only that which is within… [their] developmental level" (1978), he is describing what he calls the zone of proximal development. Problems that are in the learners' zone of proximal development are not just those problems that they cannot yet do alone; they are problems that they have the potential to one day solve by themselves. Thus, imitation leads to internalization.

While Vygotsky used this social learning process to describe young children who were learning to do simple math problems, the scope of the process is larger. The movement from imitation to internalization is the process by which, as Vygotsky puts it, learners "grow into the intellectual life of those around them" (1978). Researchers have described how knowledge is situated in the activity, context, and culture in which it is used (Collins, Brown, & Holum, 1991; Lave & Wenger, 1991). Hutchins (1995), for example, examined the role that more experienced naval navigators play in novice crewmembers' development of essential navigational skills and knowledge. The quartermaster monitors the actions of the novice watch standers as they attempt their duties, and is ready to help or take over if they are unable to complete the task to the ship's requirements. More generally, participating in the practices of a professional community, under the supervision and guidance of mentors, gives individuals access to that profession's repertoire of ways of seeing and solving problems (Lave & Wenger, 1991). Mentoring practices in professional communities are situated in the activities that the learner is attempting to master.

Many professions have institutionalized this mentor-mentee relationship in the form of a professional practicum. Schön (1983, 1987) describes how novices participate in practices they wish to learn in "simulated, partial, or protected form" under the guidance of a senior practitioner. This way of learning is suitable for learning the mores of a professional community because professionals, more than knowing basic facts and using basic skills, have a particular way of thinking (Goodwin, 1994). They make decisions based on a set of professional values, and defend those choices with profession-specific modes of argumentation and standards of evidence--that is, with a particular professional epistemology (Schön, 1983). To learn to participate in the community of practice that Lave and Wenger describe, professionals-in-training need to learn to assume an identity and values consonant with the fundamental purposes of their profession (Sullivan, 1995).

Shaffer (2006b), building on professional repertoire, communities of practice, the role of practica in professional preparation, and Vygotsky's explanation of development, introduces epistemic frame theory. "The work of creative professionals is organized around epistemic frames," Shaffer (2006b) argues: the "skills, knowledge, identities, values and epistemology that professionals use to think in innovative ways." Yet epistemic frames are not merely collections of these unrelated elements. As Shaffer (2006b) argues, "the epistemic frame of a profession is the combination—linked and interrelated—of values, knowledge, skills, epistemology, and identity" that professionals use to see and solve problems. Professionals use their epistemic frame in the context of professional action. By simulating the world of professional practice, practica provide learners with the occasion to practice professional action, and thus develop a particular epistemic frame.

In professional practica, where learners are faced with problems that are often beyond their capabilities, mentors are there to guide them. Throughout the practicum, mentors monitor learners' performance, intervening at critical moments of confusion and struggle (Schön, 1983, 1987). While in these cases the mentors may sometimes instruct in the conventional sense (Schön, 1983, 1987), they mainly function as coaches whose conversations with the learners' highlight how to navigate the obstacles of the profession. As the novices engage in the activities of the profession, the mentors reflect, and invite the learners to reflect, on the work. In this facilitation, mentors reveal to apprentices ways to go about solving problems: they model (Collins, et al., 1991). Through these reflective conversations, mentors model a way of working that requires the complex of skills, knowledge, values, identity and ways of thinking associated with the community of practice. In short, they model the epistemic frame. In modeling the epistemic frame, mentors offer learners a professional vision that they can imitate and eventually internalize.

Shaffer (2006a) describes epistemic games as simulations of professional training, and he and others have shown that those who play them develop epistemic frames (2006a, 2006b; Svarovsky & Shaffer, 2006) which can persist months after the game is over (Bagley & Shaffer, 2009). As in practica, epistemic games feature mentors who lead learners to the right way of working. In Urban Science, for example, planning consultants guide the players through a series of activities drawn directly from ethnographic study of an undergraduate planning practicum. Urban Science provides us a case study of mentor-learner interactions that can show the formation of an epistemic frame. Conversations between players and mentors in Urban Science can show the extent to which a player not only uses elements of the epistemic frame of a practice, but the extent to which the player uses elements of the frame the way a more experienced practitioner does.

In order to accurately capture players' demonstration of epistemic frames while playing epistemic games, we have developed a method to measure how game participants link the elements of the epistemic frame during game play. Epistemic Network Analysis (ENA) is a method designed to assess learner performance based on the theory of epistemic frames. This method measures whether players develop a particular epistemic frame over the course of epistemic game play (Shaffer, et al., 2009). Epistemic games are based on a theory of learning that looks not at isolated skills and knowledge, but at the way skills and knowledge are systematically linked to each other, and linked to a set of values, epistemology and identity markers. To assess a way of thinking about a professional domain means to measure a learner's formation of connections between frame elements, because this construction of a network of skills, knowledge, values, identity, and epistemology allows the learner to see and solve problems as a professional does (Shaffer, et al., 2009). ENA measures how epistemic frame elements become linked through the co-occurrence of those elements in discourse. Mapping the connections between frame elements the way that social network analysis maps connections between people, ENA assumes that the more times frame elements occur together in discourse, the more closely they are related.

This study examines how players of Urban Science develop the epistemic frame of urban planning through one particular feature of the game, namely the mentors' modeling of the epistemic frame. Specifically, we look at players' pre and post interviews and the reflective conversations between mentors and players to see whether the players imitate and internalize the epistemic frame that the mentors model. We ask three questions. First, did the players of Urban Science develop planning epistemic frames? Second, during the game, did the players imitate the epistemic frame that the in-game mentors modeled? Finally, did the players' epistemic frames during the game persist when the mentors were not present after the game?

## Methods

### Study Design

In the epistemic game Urban Science, students play the role of urban planners charged with redesigning neighborhoods in their own city. Game activities were modeled on an ethnographic study of a graduate-level planning practicum, Urban and Regional Planning 912, at the University of Wisconsin-Madison. Offered as part of a summer program called College for Kids at the University of Wisconsin-Madison, fourteen middle-school age students played Urban Science four hours a day during weekdays for four weeks during the summer of 2007. Players had no prior experience with urban planning.

The four planning consultants in the game were graduate students who underwent a one-day training in the urban planning profession, the game's activities, and mentoring strategies. The mentors met before each session to plan for the day's activities, and after each session to reflect on how the session went.

### Data Collection

Data were collected through individual interviews with each player before and after the game. The interviews were composed of questions about science, technology, and urban planning practices. Pre and postgame interviews from the game were recorded and transcribed. The questions asked were:

- What do think urban planning is?
- Do you think urban planning is important?
- What do you think it means to be a planner?
- How would you say urban planners get information for the plans they propose?
- Do planners ever work with other people?
- Do you think environmental issues are important to cities?

Data were also collected through recorded and transcribed interactions between the players and mentors during the game. Interactions consisted of one-on-one conversations where the mentor approached a player at work and asked a prepared set of questions, including:

- What are you working on?
- How is it going?
- If not going well: What have you tried? Why?
- If going well: Why is it going well?

Other interactions consisted of team meetings where the players stopped working, and met as a group to reflect on the work that they had been doing that day or during the previous activity. The mentor facilitated the meeting, asking questions that included:

- What were you working on today/before?
- How did it go?
- What did you try/How did you do it? Why?
- How would you do it differently next time? Why?

Some interactions were spontaneous conversations during game play. These interactions were initiated by players in need of some help, or by a mentor who observed that guidance was needed. The mentor usually asked the questions detailed above, but depending on the nature of the situation, the conversations varied.

## Data analysis

### Coding

Transcriptions from individual interviews were segmented into units representing one complete answer to a question, and included any follow-up questions or clarifications between the player and the interviewer. Transcriptions from in-game interactions between mentors and players were segmented into units representing one complete interaction between a player, or group of players, and mentor. A single rater coded all excerpts for elements of an urban planning epistemic frame: the interrelated set of skills, knowledge, values, identity and epistemology of the profession. The in-game interactions were coded both for the players' epistemic frames and the mentors' frames. Table 1 describes the analytic codes used in our qualitative data analysis of the in-game discourse. The analysis of the pre and post interviews used the same analytic codes.

Table 1: Analytic codes used in qualitative data analysis of in-game discourse

| Code | Description | Player Example | Mentor Example |
|---|---|---|---|
| Skills | Abilities needed to become an urban planner | Crime I managed to keep low by just not adding too much high density housing. | Are there other things you can change into housing? |
| Knowledge | Aspects of urban planning domain knowledge | One of the stakeholders wants a lot of housing…. and I'm making sure that the business, trash, and crime don't go up too high. | Anything else in terms of zoning that you think you might have to balance? |
| Identity | Feelings of belonging to a urban planning community or of being a professional | Mentor: What did they say?<br>Player 1: that our plan was sophisticated….<br>Mentor: did you feel sophisticated, presenting?<br>Player 1: more… than school | But is there a way we can look at it as planners? |
| Values | Things that are important to urban planning practice | It was pretty easy to please my stakeholders, but this plan probably wouldn't work very well for anyone else. | They [stakeholders] are all concerned about the same thing or different things? |
| Epistemology | Ways of thinking about or justifying activity within the urban planning community | That's a justification we can have for crime, that when we added the needed housing for people who work in the Schenk-Atwood neighborhood and need to live there, in order to do that, crime went up by one incident per year. | Alright, so then what would be a justification for the stakeholder who wants it way lower, what would you say to that stakeholder? |

For example, if during a conversation with a planning consultant a player mentions the value of serving stakeholders and the skill of zoning particular parcels to create a site plan that will satisfy those stakeholders, that player's excerpt is coded for values and skills, and those elements are considered linked at that moment. Similarly, if during a conversation a mentor asks a player to justify a particular zoning choice in light of a particular stakeholder's needs, that mentor's excerpt is coded for values and epistemology, and those elements are considered linked in that moment.

## Dynamic epistemic network graphs

### Epistemic adjacency matrices

For this study, the epistemic frame of urban planning is assumed to have 5 elements: skills, knowledge, identity, values, and epistemology. For any participant in Urban Science, we look at each data segment for evidence that the participant is using one or more of the elements of the urban planning epistemic frame. To construct an epistemic network from data such as this, we create an adjacency matrix for that player at that data point, recording the links between the elements of the frame for which there is evidence. For example, if in a particular segment, a player uses skills and knowledge, a link would be considered to exist between them.

By representing the epistemic frame in use during a segment as an adjacency matrix, we use the tools of network analysis to examine the cumulative impact of strips of activity on a developing epistemic frame. To construct a player's cumulative adjacency matrix, we sum the adjacency matrices of the segments of interest. We computed a series of epistemic adjacency matrices:

- one for each player's pre interview
- one for each player's post interview
- one for each player's in-game interactions, weeks 1-3
- one for each player's in-game interactions in week 3 only
- one for the players' collective in-game interactions, weeks 1-3
- one for the mentors' collective in-game interactions, weeks 1-3

### Derived Network Characteristics
Weighted density provides a measure of a frame's complexity: the overall strength of association of the frame (Shaffer, et al., 2009). We compute the weighted density of an epistemic network by calculating the square root of the sum of the squares of the associations between individual elements in the network. In order to compare weighted densities when there were a variable number of strips of activity, we normalized the data by calculating an adjusted weighted density. We calculated the adjusted weighted density by dividing the weighted density by the number of its constituent strips of activity.

Relative centrality is a measure of the relative weight of an epistemic network's constituent frame elements. By extension from the weighted density of the network as a whole, we compute the weight of a frame element from the square root of the sum of squares of its associations with its neighbors. Measuring the strength of association for a given frame element emphasizes those elements with tighter linkages to individual neighbors. We compute the relative centrality of an individual frame element by dividing its weight by the frame element with the greatest weight in the network (Shaffer, et al., 2009).

### Statistical tests
To compare the complexity of the players' epistemic frames between pre and post interviews, we calculated the weighted density of their frames based on their answers to pre and post interview questions. We then used a paired t-test to determine whether or not the weighted density of the players' frames significantly increased between the pre and postgame interviews.

To determine whether the players' in-game frames persisted after the game, we tested whether the adjusted weighted density of the players' 3rd week frames was correlated with the weighted density of their post interview frames. One player's third week consisted of only one interaction, so that player's data was removed.

## Results
Data in this section support three claims about the experience of players in the epistemic game Urban Science. First, players developed epistemic frames by playing Urban Science. Second, the players imitated the epistemic frame modeled by the in-game mentors. Third, the players' frames, as developed in the game, persisted even when the mentors are not present, after the game.

### Developing an epistemic frame
The matched pair questions players answered in post interviews contained more co-occurring frame elements than those answered in pre interviews. As a result, the weighted density of the players' post interview frames was significantly greater than that of their pre interviews (mean pre=0.1, mean post=4.4; $p<0.01$).

These changes in weighted density corresponded to a qualitative difference between players' pre and post frames. For example, one player, when asked, "What do you think it means to be a planner" in the pre interview, replied, "You sort of sketch out and you sort of visualize what will go where and how that will work out." In the post interview, the player answers the same question with considerably more detail:

> "I think it means collecting as much information as you can and it means listening to peoples' opinion and taking them into consideration. It also takes humor because you're not going to plan a place by yourself, you're going to have to collaborate with a lot of people and it takes a lot of compromising and coming up with justifications. The main goal is trying to plan and design a city and trying to improve it and making it the best you possibly can to fit the people's needs and what they want and trying to come up with a solution for all the different opinions and point of views."

Before the game, the player's answer was very vague and said generally that planners think ahead by sketching and visualizing, without any mention of the process of learning about stakeholders' needs, or working with colleagues to meet the needs of the stakeholders and the community as a whole. After the game, the player's more elaborate response refers to important elements of the epistemic frame of urban planning, like research, collaboration, compromise, and justification in the service of improving a community for its constituents.

### Imitating the modeled epistemic frame

In their conversations with players, mentors modeled the planning epistemic frame. For example, in week 1, a mentor reinforced a planning value by prompting a team of players arguing about zoning decisions to remember that their job is to represent the needs of stakeholders:

> Mentor: Are you arguing your opinion or your stakeholder's opinion?
> Player 1: My stakeholders.
> Player 2: My stakeholders' opinion. I don't agree with it, but I'm arguing for them.

The mentor's question makes clear what the correct answer is. The players' responses adopt the mentor's use of the professional term "stakeholder" and the planning value of advocating for the stakeholders' opinions as opposed to one's own whims.

The mentor then follows up with another question designed to remind the planners that they are collaborators, not adversaries:

> Mentor: My other question is: whose team are you on?
> Players 1 & 2: My stakeholders…
> Player 1: Oh, our team. Our team!

By bringing to the players' attention that they are on a team together, the mentor gets Player 1 to shift her perspective from that of a simple advocate for one stakeholder group to that of a colleague whose job it is to work with a team to serve the greater public good by satisfying all of the stakeholder groups.

In week 2, player 1, now on a different team with a different mentor, and working with a different stakeholder group, now thinks of her teammates' stakeholder groups when considering zoning changes:

> Mentor 2: ….do you feel like there are some decisions you would adjust based on what you heard here and going forward to a final plan?
> Player 1: yeah cause we gotta compromise, I can't only want these people's views to push through all the problems.
> Mentor 2: So what are some of the ones you feel you might have to adjust? Taxes was one…
> Player 1: I don't know we have a lot of things in common… Robert's group doesn't really care about greenspace, but… Cheryl's group does want more greenspace and so does mine so that's one thing we have in common and also everyone here increased housing in their plan… and I think everyone wants low crime… I don't think anyone wants high crime…

The player is now prepared to compromise because she knows that planners do not only serve one group of stakeholders. She is ready to figure out how to serve the stakeholders by finding what they want "in common."

In other words, players emphasized the same frame elements that the mentors emphasized. In terms of relative centrality, the players' frame elements followed the same sequence as those of the mentors (Figure 1):

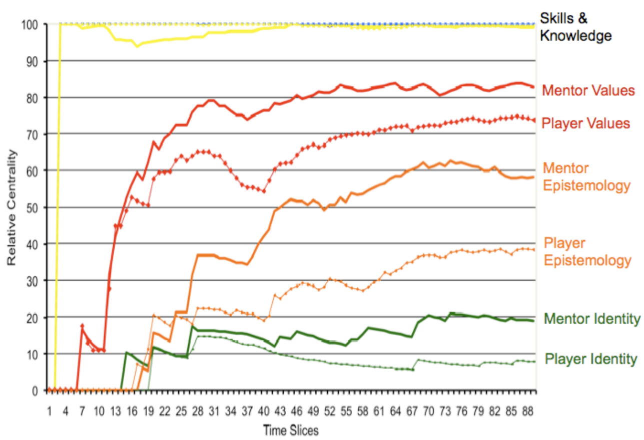



Figure 1: Collective relative centrality of players' and mentors' epistemic frames

Together, knowledge and skills were the most central for both players and mentors; as the most central frame elements, their relative centrality was always at or near 100. For both the mentors and players, the values frame element was the third most central element, the epistemology frame element was the fourth most central element, and the identity frame element was the least central element. In the conversations between mentors and players, mentors modeled a particular planning epistemic frame, and the players emulated that frame.

### Internalizing the modeled epistemic frame

Comparing players' frames as adopted in the game with their frames as demonstrated in post-interviews shows a significant relationship between them. The adjusted weighted density of the players' frames in the third week of the game correlates to the weighted density of the players' post interview frames ($R = 0.6535737$, $p = .021$).

In summary, these results suggest that a) players began to develop an urban planning epistemic frame by playing Urban Science; b) during the game, players adopted the version of the urban planning frame that was modeled by in-game mentors; and, c) the planning frame that the mentors modeled and players imitated persisted after the game when the mentors were not present.

## Discussion

This study examined whether a feature of one particular form of mentoring—the modeling of professional thinking that takes place in a professional practicum—was reproduced in the epistemic game Urban Science. Specifically, we looked at whether mentors' modeling of professional thinking contributed to players' development of the epistemic frame of urban planning. We addressed this issue in three parts.

First, we used ENA to examine the weighted density of players' epistemic frames in pre and post interviews. The players' post-interview frames were significantly denser. This change suggests that in answering the post-interview questions, players saw more connections between the urban planning epistemic frame elements than in the pre-interview. That is, the players of Urban Science began to develop a more complex urban planning epistemic frame. If, as Shaffer (2006b) suggests, professional thinking is characterized by "the combination—linked and interrelated—of values, knowledge, skills, epistemology, and identity," players of Urban Science appear to have developed a professional way of thinking by playing Urban Science.

Next, we used ENA to examine the relative centrality of individual frame elements of both the players' and mentors' frames, as enacted in their conversations together. In game conversations, the players emphasized the same frame elements as the mentors. As might be expected of more experienced practitioners, the mentors connected more frame elements more often, suggesting a more mature epistemic frame. But the ordinal position of the players' epistemic frame elements, and thus the shape of their epistemic frame, was the same as that of their mentors'. In other words, the players of Urban Science were able to imitate the mentors' professional way of talking about urban planning work. If, as Vygotsky (1978) argues, children "can imitate only that which is within…[their] developmental level," the players' successful emulation of the mentors' modeling suggests that the game's activities were within the players' zone of proximal development.

The zone of proximal development only describes the potential for internalization. In order to determine whether the mentors' modeled frame was adopted, not parroted, by the players, we compared the weighted density of the players' frames in the third week of the game with their post-interview frames. The weighted densities of these frames were correlated, suggesting that the imitated frame persisted after the game. This persistence is evidence that the players internalized professional thinking such that they no longer needed the mentors' scaffold. The urban planning professional thinking was in the players' zone of proximal development during the game, but in their actual development level when the game was finished. While it is unclear exactly when the transformation took place, the players of Urban Science began to achieve some autonomy in their ability to think as professionals, and that this autonomy was derived from their interactions with mentors. Vygotsky's hypothesis presupposes that learning processes, such as the imitation of modeled behavior, are converted into internal developmental processes. While this study does not claim to completely demonstrate the process by which children internalize external knowledge and abilities, the results presented here suggest that the imitation of modeled behavior is one step in the process of internalization. A second significant finding is that epistemic network analysis was shown to be a useful way to measure the development of epistemic frames, as well as the relationship between the players' and mentors' frames.

This study has limitations, which include a small sample with insufficient grounds for making causal claims, coarse-grained treatment of the epistemic frame, no treatment of mentoring strategies, and the use of a new and in-development method for understanding epistemic frame development. Still, the results here suggest several implications for people interested in mentoring, and in particular the ways players of games receive guidance from mentors. This study shows that mentors can model the sophisticated way that urban planners think, and that players of the game not only imitate the mentors, but develop the ability to think as urban planners themselves. In subsequent studies we are further examining the role of professional mentoring in epistemic games, and in particular, using epistemic network analysis to investigate the process by which players of epistemic games internalize epistemic frames.

## References

Bagley, E. A. S., & Shaffer, D. W. (2009). When people get in the way: Promoting civic thinking through epistemic gameplay. *International Journal of Gaming and Computer-Mediated Simulations, 1*(1), 36-52.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator, 15*(3), 6-11.

DuBois, D. L., & Silverthorn, N. (2005). Natural mentoring relationships and adolescent health: Evidence from a national study. *American Journal of Public Health, 95*(3), 518-524.

Freedman, M. (1999). *The kindness of strangers: Adult mentors, urban youth, and the new voluntarism*: Cambridge Univ Press.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, Mass: MIT press.

Kozulin, A. (2003). Psychological tools and mediated learning:Vygotsky's educational theory in cultural context. In A. Kozulin, B. Gindis, V. S. Ageyev & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 15-38): Cambridge University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*: Cambridge University Press.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press

Schaffer, R. (1996). Joint involvement episodes as context for development. In H. Daniels (Ed.), *An Introduction to Vygotsky* (pp. 251-280). London: Routledge.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*: Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*: Jossey-Bass San Francisco.

Shaffer, D. W. (2006a). Epistemic frames for epistemic games. *Computers & Education, 46*(3), 223-234.

Shaffer, D. W. (2006b). *How computer games help children learn*: Palgrave Macmillan.

Shaffer, D. W., Hatfield, D. L., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E. A. S., et al. (2009). Epistemic Network Analysis: A prototype for 21st Century assessment of learning. *The International Journal of Learning and Media, 1*(2), 33-53.

Sullivan, W. M. (1995). *Work and integrity: The crisis and promise of professionalism in America*. New York: Harpercollins.

Svarovsky, G. N., & Shaffer, D. W. (2006). *Engineering girls gone wild: developing an engineering identity in digital zoo*.

Valsiner, J., & Van der Veer, R. (1999). The encoding of distance: The concept of the "zone of proximal development" and its interpretations. In P. Lloyd & C. Fernyhough (Eds.), *Lev Vygotsky: Critical assessments* (Vol. 3, pp. 3-31). London: Routledge.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press

Wertsch, J. V. (1978). Adult-child interaction and the roots of metacognition. *Quarterly Newsletter of the Institute for Comparative Human Development, 2*(1), 15-18.

# Where to find the mind: Identifying the scale of cognitive dynamics

Luke Conlin, Department of Curriculum & Instruction, Benjamin Building, College Park, MD 20742 USA, luke.conlin@gmail.com
Ayush Gupta, Department of Physics, Toll Building, College Park, MD 20742 USA, ayush@umd.edu
David Hammer, Departments of Physics and Curriculum & Instruction, , Toll Building, College Park, MD 20742 USA, davidham@umd.edu

**Abstract:** There are ongoing divisions in the learning sciences between perspectives that treat cognition as occurring within individual minds and those that treat it as irreducibly distributed or situated in material and social contexts. We contend that accounts of individual minds as complex systems are theoretically continuous with distributed and situated cognition. On this view, the difference is a matter of the scale of the dynamics of interest, and the choice of scale can be informed by data. In this paper, we propose heuristics for empirically determining the scale of the relevant cognitive dynamics. We illustrate these heuristics in two contrasting cases, one in which the evidence supports attributing cognition to a group of students and one in which the evidence supports attributing cognition to an individual.

## Introduction

Researchers have been divided on how to answer the question "Where is the mind?" (Cobb, 1994) The "cognitivist" perspective takes the individual as the unit of theoretical and empirical focus, under the assumption that cognition is something that only brains do. While cognitivists do not ignore the social and material contexts in which cognition takes place, they analyze cognition by decomposing it into the information processing of individuals (Anderson, Reder, & Simon, 1997). That stance contrasts with situated and distributed views, which contend that cognitive processes cannot be neatly attributed to contributions of individuals when they interact with each other and their environment. Arguments for situated cognition often cite the example of a dieter solving the problem of taking three quarters of half a cup of cottage cheese by pouring out half a cup of cottage cheese, cutting it into fours, and removing one quarter (Lave, Murtaugh, & De La Rocha, 1984). A paradigmatic example of distributed cognition describes the information processing within an airplane cockpit to show that it is the cockpit as a whole—the people and instruments together as a unit—that "remembers" the safe landing speed of the airplane (Hutchins, 1995). Each challenges the attribution of cognition to an individual mind by arguing that cognition is irreducibly situated in a material and social context. The dieter's knowledge of finding ¾ of ½ a cup is inseparable from the physical materials they use to solve the problem. Likewise, no one part of the cockpit "knows" the ever-changing safe landing speed of the airplane.

There have been a number of arguments regarding the relationship between these perspectives. Cobb (1994) took them to be complementary, with one in the background of the other; Sfard (1998) also considered them complementary but "incommensurable"; Greeno (1997) took them as competing alternatives. In this article, we review an argument for theoretical continuity among these perspectives, by which they represent different scales of a complex system, rather than as fundamentally different approaches. Rather than consider the choice of perspective a matter of *a priori* theoretical commitments, we propose treating that choice as empirical. We offer heuristics for matching the choice of unit of analysis to the dynamics evident in the data, arguing that the cognitive unit can be modeled at the grain size(s) at which there are observable stabilities and coherent dynamics in cognitive activity. We then illustrate the use of these heuristics with two example analyses demonstrating empirical evidence for group and individual cognitive units, respectively.

## Manifold resources

Minsky's (1988) is probably the most widely known account of mind as comprised of manifold cognitive and metacognitive parts. He described a "society of mind" made up of a very large number of "agents." One agent he proposed was *More*. When *More* was active it was comparing two amounts and deciding which was greater. *More* was a society itself, made up of agents that corresponded to different ways of comparing different sorts of amounts, and the agents that made up more were made up of other agents. The model presented a complex system of interactions and structures at a wide range of scales. That work influenced many others' accounts, including diSessa's "knowledge in pieces" (1988) and Dennett's (1991) multiple drafts model of consciousness; it resembles and may have been influenced by "schemas" (Rumelhart, 1980).

We refer to “resources” as a generic term for the parts in such models of mind (Hammer, Elby, Scherr, & Redish, 2005). On this view, what we experience as “states of mind” correspond to stable patterns of resource activations, stability that might be only local and contextual or over extended timescales; a stable pattern of resource activations may become a resource in itself, much like Minsky’s *More*. But a stable pattern of activated fine-grained cognitive elements is not necessarily constrained to an individual mind: the pattern and its stability might extend over multiple individuals and artifacts. In this sense, the unit of “cognition” – the pattern of resources and its stability – might at times be individual or be distributed. Our claim is that empirical determination of the extent and locus of stability of such cognitive coherence should guide our sense of the unit of cognition, not *a priori* theoretical commitments.

Consistent with the views of cognition as situated and distributed, there are behaviors and cognitive dynamics that can emerge out of the interaction of multiple individuals and artifacts. Great strides have been made in the last quarter century in modeling how complex behaviors can emerge from interacting agents following surprisingly simple sets of rules. Paradigmatic examples include flocking birds (Kennedy & Eberhart, 1995), synchronized flashes of fireflies (Mirollo & Strogatz, 1990), and foraging of ant colonies (Dorigo & Stützle, 2004). Thelen and Smith (1994; Smith, 2005) have argued for complex systems-based models of cognitive development in individuals, again taking a mind to be made up of many interacting parts in dynamic interaction with each other.

Meanwhile, social science has advanced a variety of accounts of interactive dynamics, such as in conversations tending to abide by simple rules (Grice, 1989) and in shared interpretive frameworks (Goffman, 1974) that are maintained explicitly and implicitly by attending to and responding to verbal, facial and behavioral metamessages (Bateson, 1985; Tannen, 1993). These too may be understood as complex systems, at this larger scale. We suggest that these can be seen as similar dynamics at different scales, that a “society of mind” model of individual cognition is theoretical continuous with a “mind of society” model of social cognition. Thus one can model the dynamics of cognition via the synchronized activity of many cognitive resources, whether those resources are within an individual mind or distributed across minds and materials.

## What to look for: Heuristics for Data-Based Choice of Unit of Cognition

On this view, rather than as *a priori* commitment, the scale of the relevant dynamics may be determined by the evidence at hand. Our core purpose here is to propose four heuristics for making that determination: *clustering*, *persistence*, *resistance*, and *transition*. We extracted these from a range of recent work dealing with a variety of dynamics relevant to cognition including conceptions, behaviors, framing, and epistemologies (Conlin, Gupta, Scherr, & Hammer; Frank, Kanim, & Gomez, 2008; Scherr & Hammer, 2009). The heuristics are not a coding scheme, but rather a guide for empirically grounding the choice of the unit of analysis. In what follows we describe the heuristics and then apply them to two specific examples to show how they can do work in finding an appropriate unit of analysis for modeling cognition.

### Clustering

The basic idea of clustering is that a coherent cognitive state can be characterized by the ‘hanging together’ or simultaneous activation of multiple elements in the same domain or across domains. For example, how we recognize someone as greeting us is by the clustering of behavioral and speech patterns such as a hand extended for hand-shake, a smile on the face, an elevated vocal pitch, and the utterance of socially-accepted greeting words such as “Hello.” Together, these behaviors help us interpret the activity as a “greeting.” Scherr & Hammer (2009) found clusters of various kinds of behaviors (e.g., facial expressions, gestures, pitch, posture) that tended to occur synchronously within and across students working on physics worksheets in groups of four.

### Persistence

If clusters are to be suggestive of cognitive stabilities then they need to be more than just fleeting happenstance; they should persist over time. In the case of a hand-shake the behaviors might persist for a few seconds, while in other cases clusters could remain stable over much larger time scale. For example, while watching a thriller movie, behaviors such as stiff posture, intense gaze on the screen, and low perceptivity of the environment could form a cluster that persists over several minutes. In (Scherr & Hammer, 2009) stable clusters of behaviors were found to last from as little as a few seconds up to several uninterrupted minutes.

### Resistance to Change

In general, the stability of the state of a system is inferred in its ability to sustain itself in the face of perturbations. In certain moments, minor changes in the makeup of the cluster (whether the unit is the individual or the group) could challenge the coherence of physical and cognitive activity. Still, the clusters characterizing the cognitive coherence may be so stable that they can persist despite the perturbation/disturbance. Think of a greeting that lacks an important element, say, when one person extends a hand but the other person does not shake it. This can be a palpable challenge to the "greeting" frame, but it does not always result in a breakdown of the rest of the greeting activities. Fine-grained analysis of the phenomenon can illuminate how particular clusters react to naturally occurring perturbations, providing evidence on the stability and transition dynamics of clusters. More importantly, such analysis can provide empirical insight into whether the cluster (of resources, behaviors, etc.) in that moment is stabilized by an individual or a group.

### Transitions

Sudden transitions in behavior help mark the spatial and temporal boundaries of the stabilities. If there are multiple stable states in the cognitive dynamics of an individual or a group, then we should be able to see distinct transitions amongst the stable states. For this reason it is often easier to notice changes in clusters than to notice the stable clusters themselves—when something suddenly changes it highlights what was just a local stability. When a greeting is over, all of the behaviors that are characteristic of greetings end at about the same time, possibly turning into another cluster corresponding to, say, a conversation.

For an empirical example, the behavioral clusters of Scherr and Hammer (2009) were found only after noticing distinct transitions in behavior. That there were noticeable transitions led the researchers to find and clarify what the transitions were *from* and *to*, which upon further analysis were found to be coherent clusters of behaviors. The behavioral transitions are easily and reliably identified: before discussion three independent coders agreed on the timing of 90% behavioral transitions (to within five seconds). Transitions are often the result of "bids", i.e. minor changes in the clusters of behaviors. Bids that do not precipitate a transition to a new behavior cluster are evidence of the stability of the cluster. Bids that lead to a marked transition between stable clusters are also evidence of the coherence of those clusters.

In the examples that follow we will report on the clustering, resistance, and transitions of phenomena at the individual level and at the group level. We attend to the students' physical behavior (gesture, gaze, posture, etc.) as well as the substance of their discourse in applying these heuristics.

## Data Analysis

Our data is from an introductory algebra-based introductory physics course at a large public university. The students are mostly third-year life sciences majors. We draw our observations from video taken during tutorial sessions – 50-minute sessions, facilitated by teaching assistants, in which the students work in groups of four on ungraded guided-inquiry worksheets focusing on conceptual and epistemological development in physics.

First, we present an example of group-level dynamics in behavior and *epistemological framing*, that is a sense of "what is it that is going on here" with respect to knowledge (Hammer, *et al*, 2005). In the process we will show how the evidence of clustering, persistence over time, resistance to change, and spontaneous transitions all support this account in terms of group-level stabilities and dynamics. We then present a second example that supports an individual unit of analysis but not group-level cognition, in order to show a contrasting case of the scale of the unit of analysis.

More typically, we expect, the relevant scale may shift over the course of an episode, and we expect that analyzing those dynamics will be a fruitful direction for further analyses. For the present, we focus on articulating heuristics for identifying the scale of the dynamics, and for that purpose we have chosen "cleaner" examples.

## The Group as a Unit of Analysis

The first example[i] comes from a Newton's Third Law tutorial, where four students (Amanda, Bridget, Camille, and Dianna) are working together on a series of worksheet questions about a collision between a truck and a car. The particular question they are considering during this clip asks: "Suppose the truck's mass is 2000 kg while the car's mass is 1000 kg. And suppose the truck slows down by 5 m/s during the collision. Intuitively, how much speed does the car gain during the collision?" Camille is the first to offer a response, asking whether the car gains 5 m/s since the truck slows down by 5 m/s. This is an excerpt from the ensuing discussion:

| | | |
|---|---|---|
| 1 | Amanda: | The car is half as heavy, so it'll gain twice as much |
| 2 | Camille: | Ah, shoot (laughs) |
| 3 | Dianna: | Or something, Idunno. |
| 4 | Amanda: | That's what they want us to think, but this is not the real answer |

| | | |
|---|---|---|
| 5 | Bridget: | This is not the right /one/. Apparently, I think that's what they want us to say.<br><br>*…about ten seconds of silence…* |
| 6 | Dianna: | This is going...five...five meters per second, that's it's what? Acceleration or velocity? |
| 7 | Camille: | Speed. Velocity |
| 8 | Amanda: | (together with Camille) velocity. |
| 9 | Bridget: | Slows down by...?? |
| 10 | Dianna: | Velocity? |
| 11 | Camille: | Mm hmm. (pause) So the car gains ten meters per second? |
| 12 | Bridget: | I guess. |
| 13 | Dianna: | Didn't he say something about how like...something in class, like...if something's touched, the velocity, or something was changed...what was he talking about in class? Something?<br><br>*…about ten seconds of silence…* |

Camille's first intuition differed from Amanda's and Bridget's. Camille says, "the truck slows down by 5 m/s, the car speeds up by 5 m/s" but Amanda says that it would be 10 m/s because "the car is half as heavy." The discussion, however, quickly dies out and the students continue to engage in the school routine of completing the worksheet (writing best-guess answers, and occasionally seeking confirmation of vocabulary terms), until Dianna overhears the word "intuitively" from a different group's conversation, and expresses her own frustration:

| | | |
|---|---|---|
| 14 | Dianna: | I hate that word, "intuitively." |
| 15 | Camille: | See intuitively, I would think that it'd slow down, I mean speed up five meters per second. |
| 16 | Bridget: | Isn't the car...do they mean /once they hit/? |
| 17 | Amanda: | Because it is slowing down. |
| 18 | Camille: | If the if the truck if the truck if the truck, suppose the truck slows down by five meters in the collision, so if the truck is slowing down, then /I guess/ the car has to be speeding up. |
| 19 | Dianna: | The car's not moving |
| 20 | Amanda: | Yeah, it's not moving |
| 21 | Bridget: | But I think...does that mean- |
| 22 | Camille: | No, but I'm saying it says how much does does the car gain, cause it, yeah... |
| 23 | Bridget: | cause like it goes psh (gestures collision with hands) |

A vigorous discussion follows Dianne's statement as the students now focus on their own reasoning, not the worksheet; their voices and behaviors are animated as they try to communicate their ideas with gestures and explanations; and they pose original questions, such as what if the car and truck were equally heavy (not shown in transcript). During this strip of discourse, they clarify the details of the collision and the resulting speeds.

In what follows, we apply our heuristics to empirical observation of the students' behaviors and reasoning in these two segments to argue for the *group, rather than individual students* as the unit that determines the stability and dynamics of the behavioral and reasoning patterns.

## Clustering of Behavior and Reasoning: Group as Unit

As students go through the tutorial, multiple elements of their behaviors – gaze, body posture, extent of gestures, and tone of voice – are coordinated or clustered in particular patterns. Scherr and Hammer (2009) identified four stable clusters of behaviors exhibited by groups in tutorial, and these clusters account for most of the groups' time spent in tutorial (Conlin, Gupta, R. E Scherr, & D. Hammer, 2007) We will only mention the two clusters that are relevant here, *blue* and *green*.

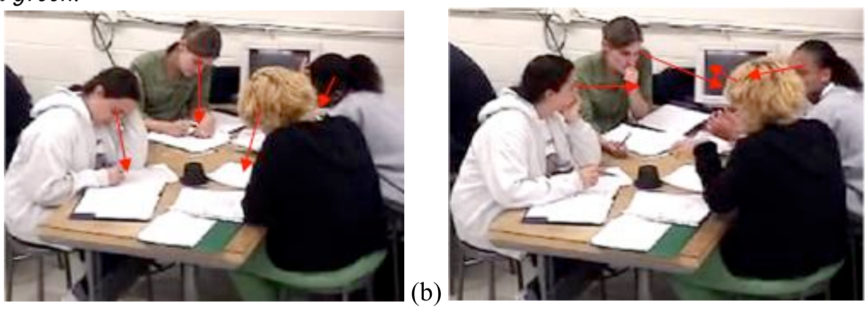
(a) (b)


Figure 1. (a) The students exhibiting the "blue" cluster of behaviors, and (b) the students exhibiting the "green" cluster.

In the first segment of the episode, the group is exhibiting the *blue* behavioral cluster (see Figure 1a) -characterized by downward gaze, hands writing or resting on the table, scant and mostly deictic gesturing, low or flat voices, and intermittent speech with no overlap. In the second segment, the students transition *as a group* to the *green* behavioral cluster (Figure1b), in which they tend to make eye contact, sit up, gesture, and speak in animated voices with overlapping speech. The arrows in Figure 1 show how the gaze of the students is coordinated – and the locus of their gaze shifts in synchronous patterns making it meaningful to talk about the *group* as gazing towards the worksheet during the earlier segment and towards one another during the second segment.

These behavioral clusters also often indicate locally stable ways in which the students were framing the activity, either as *completing the worksheet* or as *having a discussion* (Scherr & Hammer, 2009). At first, their speech is in the service of finding the answer to the worksheet question. Amanda, Dianne, and Bridget all but say this explicitly by suggesting that Camille write down an answer that "they want us to think" rather than one that makes sense to her. Furthermore, their answers are brief and only weakly supported with justification. Dianna tries to settle the vocabulary ("five meters per second, that's its what, acceleration or velocity?") and tries to remember the information the professor gave in class about collisions. The content of their speech taken in conjunction with their behaviors suggests that the group is initially framing the activity as *completing the worksheet*. After Dianna expresses her frustration with the word "intuitively," the group starts to clarify and piece together the mechanism of what happens during a collision, rather than simply comparing the speed before and after. Their speech and behaviors are clustered around *having a discussion* about a physical phenomenon: Bridget, Amanda, and Dianne try to clarify the conditions of the collision; Camille reasons about the reactions of the truck and car as related, "if the truck *is slowing down*, then /I guess/ the car *has to be speeding up*" (emphasis added); Bridget's gestures try to simulate the motion of the car.

## Transition in Behavior and Reasoning: Group as Unit

There is a sharp transition in behavior at both the individual and the group levels that starts with Dianna's commentary that she "hates that word, 'intuitively.'" Her affective stance shook the group away from the worksheet, opening up the space for a more personal discussion. Everyone looks up to pay attention to Dianna – a sharing among friends. To Camille, this provided the space to clarify her original intuition, an action she did not take in the first part of the episode when the focus was on figuring out what "they want us to think." Camille's emphasis on the word 'intuitively' and tone support this interpretation. Next, Camille emphasizes the compensation of speeds between the truck and the car; Amanda gestures the speeding up of the car after the collision. The emotional outburst of Dianna cued them to think about and communicate their personal opinions, leading to a cascade of utterances towards mechanistic pursuit of the phenomenon. The coordinated nature of the transition in behavior (from *blue* cluster to the *green* cluster) and framing (from *doing the worksheet* to *having a discussion*) makes it appropriate to think about the *group* as transitioning rather than four individuals separately making these transitions. This supports the theoretical coupling of framing and behavior, since behaviors act as the metamessages that comprise and maintain the group's framing (Bateson, 1985).

### Resistance of Behavior and Reasoning: Group as Unit

*Bids to switch frames* – instances in which one students' behaviors and/or reasoning stand in contrast to that of the established coherent frame (Scherr & Hammer, 2009) – can provide valuable insight into the dynamics and appropriate scale for a cognitive unit. In this episode, while the group is exhibiting the blue behavioral cluster engaged in *completing the worksheet*, Dianna makes a bid to have a discussion (turn 13). This is evident through her behavior, since she sits up, speaks more loudly, and has much more pronounced prosody when asking, "Didn't they say something about how like...what was he talking about in class?" One by one, the other students briefly look up as she is talking. As she trails off, however, the other students return to their blue cluster behaviors one by one. There is no talking for 10 seconds as the group goes back to completing the worksheet. Dianna's bid to change behaviors was not taken up.

Epistemologically, Dianna's question is a bid to try to remember what the professor said in class, a departure from the group's engagement in finding the worksheet-directed intuitive answer to the gain in the car's speed. That the rest of the group simply turns back down to their worksheets without responding to her question sends the message that they did not see this as a productive line of pursuit. Dianne's deviation from the shared cluster of behaviors and framing could not disrupt it, rather it was Dianne who returned to that shared activity. This illustrates the stability of the cluster and its extension beyond single individuals.

### Persistence of Behavior and Reasoning: Group as Unit

Additional evidence for the *group* being the appropriate unit of analysis here comes from examining what keeps the behavior or framing stable during these segments. We find that the behaviors serving as meta-messages communicating the fine-grained stances students are taking in their utterances. When a student makes animated gestures or talks loudly, it sends the message of a discussion or argument in contrast to hunching over your personal worksheet. Seeing one member clarifying an idea might prompt a second student to add to it, which supports and encourages the effort of the first creating feedback loops that help in the persistence of the activity. The moment-to-moment behaviors and utterances establish a shared understanding of the nature of the activity – that is self-sustaining. Before the transition, individual behaviors such as hunching over the worksheet, exchanging short vocabulary or quick questions in low voices, reinforces the *group's* shared sense of the activity of filling the worksheet; and later, after the transition, their pursuit of a mechanistic explanation of what happens in the collision is built out of bits of contribution from the different members. In either segment, the persistence of behaviors and reasoning is a group endeavor, even though individuals are enacting these behaviors and utterances.

## The Individual as a Unit of Analysis

The second example[ii] comes from a tutorial on shadows and light. In this tutorial, students try to build a model of light by exploring the patterns and motions of light created by a light bulb and an aperture. The question being considered by the group during the clip is, "Why does the light on the screen moves to the left when the bulb is moved to the right?" Veronica and Jan have a disagreement over the mechanism of light shining through the aperture. Here is a piece of the conversation:

| | |
|---|---|
| Jan: | All the rays are going like this. So, it's kind of like polarizing it. |
| Veronica: | Mmm, not really. |
| Jan: | You sure? |
| Veronica: | It's just, well, it's just, guys, you're making it- you're trying to make it too difficult. It's just, the light goes out. It only goes through that one circle. So, obviously, if it is down here, and I'm looking through that circle. Look, you're sitting down here. You're looking at this big cardboard. You're looking up through that little circle. All you're going to see is what's up there. It's a direct line. |
| Jan: | Look, I see what you are saying, alright? But, I'm just trying to make it like physics-, physics-oriented. |
| Veronica: | It *is* physics-oriented, that's just the way it is. |
| Jan: | Okay. |

### Clustering of Behavior and Reasoning: Individuals as Unit

Throughout the conversation presented as well as before and after it, Veronica is sitting up, looking at the other group members, speaking in a loud, animated tone, and gesturing (Fig 2). Jan, who is sitting across the table from Veronica, for large parts of the tutorial, is hunched over, and drawing on the tutorial worksheet. During the moments of the conversation above, she does look up at Veronica and speak in a clear and loud voice, but these were fleeting moments. The larger pattern was of students conducting themselves differently. This lack of synchronicity and coordination among the behaviors suggests a clustering at the individual level in this case, not at the group level.

Figure 2. Veronica is gesturing as Jan is writing on her worksheet.

Again, the contrasting behaviors indicate that Jan and Veronica are framing the activity differently, also evidenced by their very different epistemological stances. Veronica is appealing to her intuitive sense of mechanism of how light travels, while Jan tries to make what Veronica is saying more "physics-oriented," by using technical terminology. Veronica objects with a different take on what it means to be physics-oriented, namely that it is "just the way it is." For Veronica, the activity is about making sense, while Jan is trying to get the formal answer.

### Persistence and Resistance of Behaviors and Reasoning: Individuals as Unit

Both Veronica and Jan activate a stable set of epistemological resources throughout this clip. In fact, this stability persists on much larger timescales than one tutorial period. Lising & Elby (2005) analyze Jan's work throughout the course and argue that her epistemological stance, in which there is a barrier between formal and everyday reasoning, is relatively stable and robust throughout the semester.

The persistence of each student in their epistemological framing is stable in spite of direct challenges from each other. Veronica expresses impatience with Jan by bluntly correcting her with, "Mmmm, not really." She criticizes Jan for "trying to make it to difficult," and proceeds to explain the phenomenon with common-sense reasoning. Jan contents that she is just trying to make it "physics-oriented," with the implication that Veronica's common-sense explanation is not physics-oriented. Veronica defends herself from this challenge: "It *is* physics-oriented. It's just the way it is." When Jan responds with, "Okay," the group goes silent. It appears that they just agree to disagree, persisting in their epistemological framing despite the explicit bids made to change them.

### Transition in Behaviors and Reasoning: Individuals as Unit

Just as there is no coherent clustering at the group level, there are no clear transitions at the group level between clusters. This seems to be the exception rather than the norm, since most of the time groups spend in tutorial is in one of four behavioral clusters separated by sharp transitions (Scherr & Hammer, 2009). This counts as evidence *against* the group being the unit of cognitive analysis during this clip. This lack of group transitions, taken in conjunction with the other heuristics above, constitutes a strong empirical case for taking the individual as the unit of analysis here.

## Conclusions

Our argument is two-tiered. On one tier, we analyze the behavior and discourse within two episodes of student group work in introductory physics tutorials. We found evidence for stabilities and dynamics of cognition at the group level and the individual level, using four empirical heuristics for identifying the scale of the cognitive unit. In the first example we demonstrated that the clustering, persistence, resistance, and transition reflect dynamics at the level of the group as a whole. It is the collective group frame that enables them to reason about the mechanisms involved in a collision in a more sophisticated way. In the second example we demonstrated how clustering, persistence, and resistance resided not with the group but rather on the level of individual students' epistemological stances and their conceptual understanding of how light travels. Veronica and Jan held different views not only of light but also of physics, and these views remained stable in the face of their mutual bids to change the other's mind.

One tier up, we are arguing that these analyses demonstrate how the choice of cognitive unit can be made empirically by attending to the dynamics and stabilities evident in the data. This is in contrast to the positions on both sides of the ongoing debate about the unit of analysis, with cognitivist and situativists basing their decision largely on theoretical considerations. We introduced four empirical heuristics for identifying the scale of cognitive dynamics. Our analyses of the two episodes highlight the utility and generativity of our empirical heuristics, although we do no claim that these span the set of evidentiary supports for determining the cognitive unit. All of this is underwritten by the resources and framing model of mind, since resources can be distributed within an individual mind or across several individuals and artifacts. Hopefully our approach will enable greater communication and collaboration between cognitivists and those who model cognition as situated and/or distributed.

## Acknowledgments

We would like to express our thanks to the National Science Foundation for financial support, and to the Physics Education Research Group at the University of Maryland, College Park, for helpful discussions.

## References

Anderson, J. R., Reder, L. M., & Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational Researcher, 26*(1), 18-21.

Bateson, G. (1985). A Theory of Play and Fantasy. *Semiotics: An Introductory Anthology*.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7), 13.

Conlin, L. D., Gupta, A., Scherr, R. E., & Hammer, D. The Dynamics of Students' Behaviors and Reasoning during Collaborative Physics Tutorial Sessions. *Red, 24*(15), 4.

Dennett, D. (1991). Consciousness Explained. LOCATION: the Penguin Press.

diSessa, A. (1988). Knowledge in pieces. *Constructivism in the computer age*, 49-70.

Dorigo, M., & Stützle, T. (2004). *Ant colony optimization*: the MIT Press.

Frank, B. W., Kanim, S. E., & Gomez, L. S. (2008). Accounting for variability in student responses to motion questions. *Physical Review Special Topics-Physics Education Research, 4*(2), 20102.

Goffman, E. (1974). *Frame analysis : an essay on the organization of experience*. Cambridge, Mass.: Harvard University Press.

Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5.

Grice, P. (1989). *Studies in the Way of Words*: Harvard Univ Pr.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. *Transfer of learning: Research and perspectives. Greenwich, CT: Information Age Publishing*.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science: A Multidisciplinary Journal, 19*(3), 265-288.

Kennedy, J., & Eberhart, R. (1995). *Particle swarm optimization*.

Lave, J., Murtaugh, M., & De La Rocha, O. (1984). The dialectic of arithmetic in grocery shopping. *Everyday cognition: Its development in social context*, 67–94.

Minsky, M. (1988). *The society of mind*: Simon & Schuster.

Mirollo, R., & Strogatz, S. (1990). Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics, 50*(6), 1645-1662.

Scherr, R. E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction, 27*(2), 147.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational researcher, 27*(2), 4.

Tannen, D. (1993). *Framing in discourse*. New York: Oxford University Press.

Thelen, E., & Smith, L. B. (1994). A dynamic systems approach to the development of cognition and action. Cambridge, Mass.: MIT Press.

---

[i] The Newton's 3rd law clip is also discussed in Scherr & Hammer (2009).

[ii] The Shadows and Light clip is also discussed in (Lising & Elby, 2005).

# Representational Technology For Learning Mathematics: An Investigation of Teaching Practices in Latino/a Classrooms

Phil Vahey, Teresa Lara-Meloy, SRI International, 333 Ravenswood Dr. Menlo Park CA
philip.vahey@sri.com, teresa.lara-meloy@sri.com
Judit Moschkovich, Griselda Velazquez, University of California Santa Cruz, Education Department, 1156 High Street, Santa Cruz CA, jmoschko@ucsc.edu, gvelazqu@ucsc.edu

**Abstract:** This paper explores representation- and language-rich mathematics instruction in two classrooms with low-income Latino/a students. The two classroom teachers were part of a larger study investigating the use of SimCalc in middle school, and their classrooms had learning gains greater than the mean gains for the study overall. Prior analyses documented that these two teachers implemented exemplary classroom discourse practices: they engaged students in discussions with high intellectual work and showed high levels of responsiveness to student statements (Pierson, 2008). Such practices are consistent with both the research on representationally-rich mathematics environments, and instructional practices for students from non-dominant communities. We use transcript excerpts to illustrate these teachers' practices in detail. We found that these two teachers enacted some similar practices, but they also had different teaching styles, different approaches to connecting representations, and different ways of supporting mathematical discussions.

## Introduction and Significance

As the population of Latino/a students increases in U.S. schools, so do concerns with their needs in mathematics classrooms. One concern is performance on assessments. Only 13% of Latino/a students scored at or above proficient on the 2005 8th grade NAEP exam, compared to 39% of white students and 47% of Asian students (Gándara & Contreras, 2009). Many schools attempt to address these concerns by placing large numbers of Latino/a students in remedial courses designed to address the needs of students who are "at risk," a decision which may exacerbate the situation. Research shows that the problem lies not in the students but in their access to quality teaching. In general, low-income Latino/a students attend racially segregated poor schools and have little access to the kind of teaching that supports the development of conceptual understanding or academic discourse (Gándara & Contreras, 2009). However, there are examples of classrooms where this is not the case. Quantitative data from the SimCalc study (Vahey, Lara-Meloy, Knudsen, 2009) and qualitative research with this student population has shown that Latino/a students can learn complex mathematics and participate in mathematical discussions (i.e. Moschkovich, 2002) when they have access to quality teaching.

In this paper we examine the teaching practices of two teachers, providing a sharp contrast to the perspective that Latino/a students need 'remedial" teaching. We analyze two instances of mathematics instruction that support conceptual understanding and mathematical discussion as students participate in representation-rich mathematics environments. There is reason to believe that there is consonance between the literature on improving instruction for low-income students from non-dominant linguistic backgrounds and the literature on the use of representationally rich technologies in mathematics (Vahey, Lara-Meloy, & Knudsen, 2009). Both highlight the use of multiple ways to represent ideas, point to supporting students as they make connections among these multiple representations, and point to the importance of language rich practices. However, there has not yet been a systematic attempt to integrate the representational affordances of technology with research on improving mathematics instruction for students from non-dominant linguistic backgrounds. This leaves educators with little guidance on how to use technology for this underserved student population.

This paper begins to address this gap by using two lenses to describe the instructional practices of two teachers who were successful in using technology to support students in their primarily high-poverty Latino/a classrooms in learning important conceptual mathematics. We explore the ways in which these successful teachers engaged in practices that were consistent with (or at odds with) research-based recommendations for mathematics instruction that is (a) representation-rich, and (b) supports mathematical discussions.

## Background

We ground our exploration in the Scaling Up SimCalc study (Roschelle et al., in press) for three key reasons. First, SimCalc is based on a rich theoretical and empirical base, and so provides an excellent proxy for representationally rich technology use more generally. Second, the Scaling Up SimCalc study was a large scale randomized controlled experiment that not only found significant results across a wide range of student demographics, but also collected a rich array of data including many videotaped lessons. Finally, the study included large numbers of Latino/a(1) students (about half of the 1621 total students) from poor schools, many in the Rio Grande Valley, which borders Mexico and is one of the poorest regions in the United States (the study denotes this area as Region 1, as it is served by the Region One Education Service Center). The SimCalc study is an ideal place to begin our exploration into the connections between research on representationally rich technologies and instructional recommendations for teaching students from non-dominant language groups.

## The Scaling Up SimCalc Study

For over fifteen years the SimCalc project has had the goal of ensuring that all learners have access to complex and important mathematics (Kaput, 1994). To achieve this goal SimCalc places motion phenomena at the center of learning (see Figure 1), enabling students to build on existing cognitive and social competencies. Students in our studies, including traditionally low-achieving students, construct rich stories about motion over time, and use narratives as a resource for interpreting graphical and tabular representations of motion. SimCalc also allows students to play and replay a motion simulation as many times as they wish, providing more students access to these fundamental resources than is possible using traditional static media.

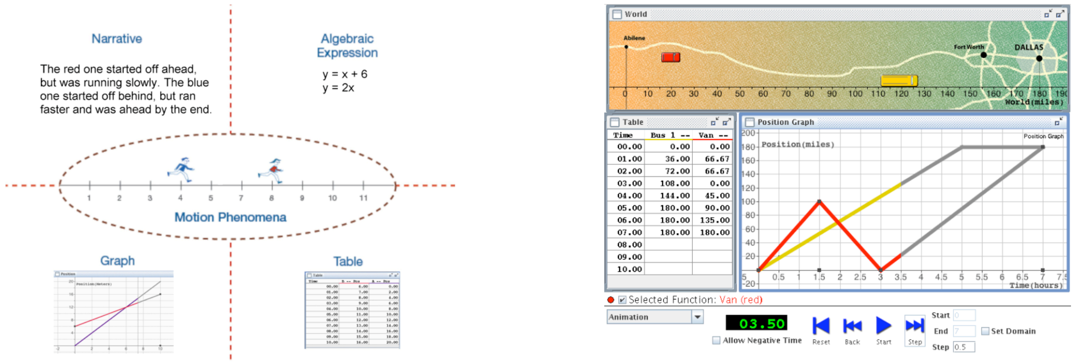



Figure 1. SimCalc linked representations (left); SimCalc MathWorlds® activity screenshot (right)

This engagement with motion leads to the study of functions through linked simulations, graphs, tables, and symbols. Students engaged with SimCalc can directly manipulate a mathematical representation such as a graph, and immediately see the effects on other linked representations (Roschelle et al., 2000). Formal mathematical vocabulary and symbolic forms can then be introduced *after* students have experiences with motion, narratives, tables, and graphs. In this way the vocabulary and symbols are *about* something, and can be understood as a compact and precise way of describing phenomena. The features of direct manipulation, multiple representations, and experience-before-formality may be particularly beneficial to students from non-dominant linguistic backgrounds, as these features alleviate some of the language demands found in mathematics classes, allowing students to work directly with mathematical objects (Vahey, et al., 2009).

These same features can allow students to more fully participate in classroom discussions. The linked representations provide a shared set of referents for students and teachers to explore: they can replay a motion or make changes in one representation to see the changes in the others. Students have opportunities to use a wider range of verbal and nonverbal communication acts, such as pointing: "See, right *here* the boy starts running faster." Students also have opportunities to use academic mathematical language for a communicative goal (e.g., Does *going longer* refer to *time* or *distance*?). This multi-faceted approach is consistent with recommendations for supporting mathematical discourse and developing vocabulary (Moschkovich, 2007a, 2007b) and it contrasts with traditional approaches to teaching academic language that rely on drill and practice or memorization.

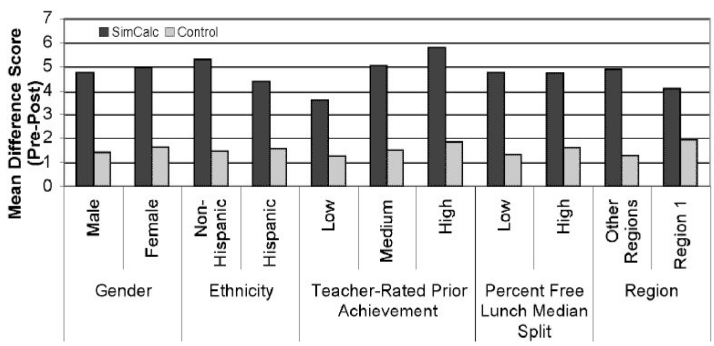



Figure 2. Mean student-learning gains by subpopulation group.

The Scaling Up SimCalc study found the SimCalc approach to be successful in meeting the needs of a variety of teachers and their students. Ninety-five seventh grade teachers across varying regions in Texas participated in a randomized controlled experiment in which they implemented a SimCalc-based three-week replacement unit. An analysis of the results showed a large and significant main effect, with an effect size of 0.8

(Roschelle et al., in press). This effect was robust across a diverse set of student demographics including gender, ethnicity, teacher-rated prior achievement, and poverty level (measured by the percentage of the campus eligible for free or reduced price lunch) (Figure 2). Consistent with our other data, we see that the students in Region 1 (the Rio Grande Valley) who used SimCalc had greater learning gains than students in the control condition, and these learning gains included improvement on items that target conceptual understanding.

## Theoretical Framing

We take a socio-constructivist theoretical approach. We assume that students learn through interacting with materials and technological artifacts while participating in discourse, that mathematical dscourse is multi-modal and multi-semiotic (O'Hallaran, 2000), and that it includes not only talk, but also other tools such as inscriptions or animations and several modes such as oral, written, and gestural.

### Recommended instructional practices for all students

Research in mathematics education describes teaching that promotes conceptual development as having two central features: one is that teachers and students attend explicitly to concepts, and the other is that students wrestle with important mathematics (Hiebert & Grouws, 2007). Teachers face a considerable challenge in balancing both of these features in their teaching. In particular, teachers working with Latinos/as often focus on procedural rather than conceptual aspects of mathematics (Gándara & Contreras, 2009). In contrast, effective environments for students from non-dominant linguistic backgrounds should provide "abundant and diverse opportunities for speaking, listening, reading, and writing" and "encourage students to take risks, construct meaning, and seek reinterpretations of knowledge" (Garcia and Gonzalez, 1995, p. 424). These practices require that teachers view language as a resource, not a deficiency (Gándara & Contreras, 2009), and reject models of their students as intellectually disadvantaged (Garcia and Gonzalez, 1995).

Framing research on students from non-dominant linguistic backgrounds is complex. This student population is heterogeneous (e.g. in some geographical areas this population includes students from many countries who speak many languages). Within our relatively constrained population (students from two classes in the Rio Grand Valley) almost all students are labeled as Hispanic. Although we do not have an exact account of how many students in these classrooms were learning English, it is likely that students in these two SimCalc classrooms, like students in this geographic area in general, vary along a spectrum of monolingual to bilingual. We assume that some students in these classrooms may be learning English; some students may be first generation immigrants, and some families may have lived in Texas for several generations. With that complexity in mind, we examine teaching practices for these two teachers because we believe they are illustrative of teaching strategies that support gains in conceptual mathematics for Latino/a learners, some of whom are likely to be learning English.

#### <u>Use of representations</u>

A key feature of mathematics is the use of representations, not only to overcome the limitations of human memory (as in the use of simple lists), but also to embed computational rules into symbol systems (e.g. algebraic notation), and to re-represent complex relationships in ways that can be more easily perceived by those who have facility with the representational system (e.g. the graph of a function) (Ainsworth, 2006). The use of representations is both a requirement for full participation in everyday quantitative reasoning (e.g. it is necessary to interpret graphs commonly shown in daily newspapers), and an important tool that can allow students to engage with mathematical objects as they also grow in their mathematical skills and understanding (Kaput, 1994). Representations, and dynamic representations in particular, can be particularly powerful learning tools when they are part of mathematical discourse: they can support shared focus of attention, allow gestural and physical communication to supplement verbal communication, and provide meaningful feedback that is consistent with the mathematical phenomena under investigation (Moschkovich, 2008; Roschelle et al., 2000).

Recommendations for instruction for students from non-dominant linguistic backgrounds (e.g. August and Shanahan, 2006; Echevarria et al., 2004) also include the use of representations. These recommendations focus on the use of visual artifacts to (a) offload demands on language, and (b) represent abstract ideas using illustrations or objects (e.g. holding up an illustration of a triangle when referring to a triangle in a geometry lesson), rather than as a way to provide insight into complex mathematical concepts.

#### <u>Use of mathematical discussions</u>

One way to engage students in both attending to and wrestling with important mathematics is through mathematical discussions. It is generally accepted that "mathematical discourse" and "academic language" are important for all students to develop, and are especially important for students from non-dominant communities. As students participate in these activities they are learning to communicate mathematically by making conjectures, presenting explanations, constructing arguments, and so on. When describing mathematical discourse we should not confuse "mathematical" with "formal" or "textbook." Textbook definitions and formal ways of talking are only one aspect of school mathematical discourse. It is also important to avoid construing everyday and academic mathematical discourse as opposites (Moschkovich, 2007a), as some everyday experiences may provide resources for communicating mathematically.

Some recommendations for teaching reduce mathematical discourse to addressing vocabulary through direct instruction, drill and practice, or memorization. In contrast, other recommendations are based on research that shows that academic discourse is more than vocabulary, and that vocabulary is most successfully learned through instructional environments that are language rich, actively involve students in using language, require both receptive and expressive understanding, and require students to use words in multiple ways over extended periods of time (Blachowicz & Fisher, 2000). Additionally, for mathematics instruction, learning to communicate mathematically is not primarily a matter of learning vocabulary, as students also need to develop conceptual understanding and learn to describe patterns, make generalizations, and use representations to support their claims (Moschkovich, 2002 and 2007b).

## Analysis

In our exploration we examine and summarize the teaching strategies in one lesson selected from the classes of each of two teachers from Region 1, Teacher E and Teacher M. We chose these classes because these students are commonly labeled "at-risk" for poor mathematical performance: one class is 96% Latino/a (Teacher E) and the other is 100% Latino/a (Teacher M); and both classes are in schools where over 84% of the students are eligible for free or reduced price lunch. However, a previous quantitative study (Pierson, 2008) found that students in these two classrooms had gains greater than the mean gains for all classes using SimCalc. Pierson's analysis indicates that one possible reason for these gains may have been that these two teachers engaged their classes in discussions with high intellectual work, and showed high levels of responsiveness to student statements (Pierson, 2008). Intellectual work reflects the cognitive work requested from students with a given turn of talk. Responsiveness is an attempt to understand what another is thinking, displayed in how the teacher builds, questions, clarifies, takes up or probes that which another says.

Pierson (2008) shows that, for the 13 classes that had video data available, these two types of teacher moves were correlated with student gains. The two selected teachers stand in contrast with another teacher from a nearby school with a similar student population (95% Latino/a students and 93% students on campus on free and reduced lunch) who had lower than average intellectual work, responsiveness, and student gains. Our question moving forward is: in what ways did these two teachers, selected for exemplary mathematical discourse practices, engage in practices that were consistent (or at odds) with recommended instructional practices for the use of representationally rich mathematics environments and supporting mathematical discourse? Given the correlational findings in Pierson, our goal here is to provide a more detailed qualitative description of interactions in these two classrooms. These two descriptive cases ground both research and practice: while the correlational study showed *that* the quality of these two teachers' discourse was related to student gains, our goal here is to describe *how* these two teachers enacted these practices during a lesson.

The data for this study are records of teacher activity at the front of the classroom over a single lesson. Thus, our analyses will focus on teacher's uses of talk and inscriptions. Our data will not let us investigate non-observable features such as teacher beliefs, or those features that take an extended time to develop, such as the initial setting of classroom discourse norms. The classroom videos and transcripts we analyzed for teachers E and M show the same lesson from the SimCalc 7th grade curriculum, entitled "On the Road." During this lesson, students are expected to write stories explaining the motion of a bus and a van over several trips from Abilene to Dallas. Students must make sense of piecewise linear functions (see Figures 3 and 4) of increasing complexity. Students are asked to interpret functions that consist of multiple slope lines, including some with zero and negative slopes. In the first situation, students observe a position graph and corresponding computerized simulation of the motion of a bus and a van, and then write an explanatory story. In the second situation, students are asked to predict the motion of the bus and the van based on the graph before viewing the simulation to verify their predictions.

In our analysis we first partitioned the classroom videos into a set of curricular segments. These segments are based on curricular topics, as dictated by the teacher (and, of course, heavily influenced by the SimCalc materials). Within each segment we investigated episodes in which the participants focused on a particular mathematical topic. While the materials influence these topics, episodes were typically driven by the particular classroom interactions, and we expect episodes to vary significantly by classroom. In this paper we focus on representative classroom episodes that illustrate how these teachers (a) made significant use of representations or (b) engaged students in significant mathematical classroom discourse.

Our analysis shows that Ms. M and Ms. E used various strategies during these lessons. Since some of the teaching strategies were similar across the teachers and others differed, we emphasize that no single aspect of the teaching accounts for student gains, that there is no single way to enact best practices, and that there is no simple formula for teaching this student population. In fact, a strength of this analysis is to show that teachers with different teaching styles can instantiate best practices in such a way that students commonly labeled "at risk" can and do learn complex and conceptual mathematics. Additionally, although we present examples of how the teacher used representations and supported discussion separately, these two are dialectically related, and this separation is only for the purpose of our illustration.

### Teaching Practices that Connect Representations

As the teachers led class discussions based on position graphs, we note that both teachers consistently pushed the students to make connections between the two functions represented on each graph, and across multiple representations (simulation, narrative, and graph). In addition, after several attempts to have the students describe these connections, the teacher would then explicitly state the connection she considered most salient to the activity. Both teachers did this on occasions where the students had already described the desired connection, as well as cases in which the students had not, presumably so that all students had some exposure to the desired connection.

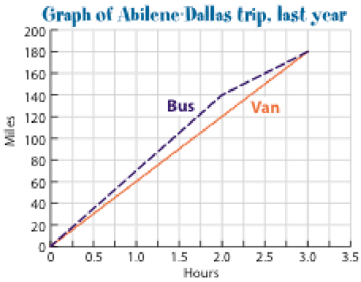



Figure 3. Last year's graph of the Abilene-Dallas trip.

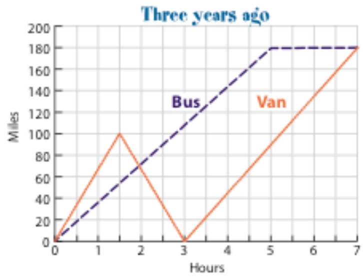



Figure 4. The trip from three years ago.

### Example 1: Ms. E (see Figure 3)

Ms. E conducted this class is in a computer lab, with a 1:1 student to computer ratio. Students were seated facing the computers, lined up along the wall. The teacher had a document camera projecting her workbook in the center of the room. During whole group discussions, she often pointed or traced her pencil over the graphs or text in her workbook projected via the document camera. This type of gesturing is a common strategy for helping students connect what is being said with the activity tasks. The following episode took place early in the lesson. The students worked in pairs at computers as they viewed the simulation that is represented by the graph in Figure 3. As this episode begins, the teacher was orienting the students to a whole-class discussion.

T: Ok. Did everybody get a chance to go through the simulation?
Ss: Yes.
T: It says: (Reading from the workbook) "What information can you get about last year's trip from looking at and analyzing the graph? Write down everything you can think of and be sure to include the speeds of the vehicles." Ok. So the first thing that we noticed is that we have two graphs. One represents who and who?
Ss: The bus and the van.
T: Ok. Very good. The bus and the van. Ok. What can you tell me by looking at the graph?
Yuri: The bus was going faster --
Amy: -- The bus was going faster but then it slowed down and the van caught up.
T: Ok. Yuri and Amy said that the bus was going faster but then what happened Amy?
Amy: The bus slowed down and the van caught up.
T: Ok. [Writing in her workbook] "The bus was going faster." So let's write that down.

In this excerpt we see the teacher explicitly asking the students to connect the graph to the simulation they had just viewed ("One represents who and who?", "What can you tell me by looking at the graph?"). However, it is not obvious (to the teacher or to the observer) whether the students were relying on their recollection of the simulation, the markings on the graph, or both. For example, the teacher asked for information that could be derived from the graph, but the source of Yuri and Amy's response is unclear, as they could have been referring to the simulation. Although the teacher accepted these responses in this interaction, she followed this conversation with a detailed analysis of the speeds of the bus and the van, using statements that connected the information they had gathered from the simulation with the details in the graph. This is shown in Example 3, in which we analyze how the teacher leveraged student language as a resource for not only connecting representations, but also for supporting conceptual understanding and language development.

### Example 2: Ms. M (See Figure 3)

Ms. M's class was also held in a computer lab, with one student at each computer facing the front of the class. The relevant SimCalc software file was displayed for the class with a projector. Ms. M seemed to have a good rapport with her class, often teasing them and being willing to entertain their outlandish stories. In this example we see Ms. M asking students to read their stories to the class (as in Example 1, the class had already viewed the corresponding motion simulation). Ms. M challenged Dan's story not by questioning the (possibly inappropriate) content, but by questioning the relationship between the story and the graph.

| | |
|---|---|
| T: | Everybody listen to Dan's story. Go ahead Dan. |
| Dan: | Alright. The bus driver is going 70 miles per hour. But then a monster got in the way and the bus driver had to slow down to 40 miles per hour. And, then he shot him with an emergency shotgun, (student laughter) but not before the monster ate him. But at the end they played in memory of him. But I had to take over and drive the bus-- |
| T: | -- But the bus -- was already there, if I'm interpreting correctly, you already said that they got there and that's when the monster attacked him? |
| Dan: | No! [unintelligible] |
| T: | Read it again. Go ahead. |
| Dan: | "The bus was going 70 miles per hour. But then all of a sudden, a monster gets in the way (he makes a monster growl)." |
| T: | And it slowed down? |
| Dan: | Yeah. And the bus, it slowed down, he got scared, and took the emergency shotgun and shot him but not- not- |
| T: | While he was driving? |
| Dan: | Yeah, while he was- like this (pretends to shoot a gun, students laugh). |
| T: | Okay. That's kind of scary. |
| Dan: | But not before the monster ate him. So I had to take over and drive the bus. And the game is played in memory of him (reads imagined name and date of birth and death). |
| T: | But okay, in the real world, okay, if they eat something, how is it staying constant? (gestures to the graph projected on the board) How does it stay constant (gestures to the graph of the bus)? 40 miles per hour. Unless you were there while he was being eaten, and you put your foot, and you stayed steady. Understand? |
| Dan: | Yeah |
| T: | So maybe if you would have arrived in Dallas (points to Dan) and then the monster could have eaten the bus driver. |

In this excerpt we see the teacher challenging Dan to describe exactly how his story related to the graph. She points out that in Dan's story the driver is eaten en route, without the bus stopping She pushes Dan to make the connection between his story and the fact that the graph shows the bus slowing from a constant speed of 70 mph to another constant speed of 40 mph. While she was apparently willing to accept that the monster scaring the driver would result in the bus slowing down in a manner consistent with the graph ("And it slowed down?...Okay"), she was not willing to accept that the monster could then eat the driver with the bus staying at the constant speed of 40 mph. Instead she provided an alternative narrative ("So maybe if you would have arrived in Dallas and then the monster could have eaten the bus driver").

In sum, we see that both teachers leverage the simulation capabilities of SimCalc to provide a shared context of motion that can be referred to by all students. They also used SimCalc's position graphs to provide a shared mathematical representation that grounded the discussion, and a set of stories that were used to explain the graphs and simulations. The teachers then spent considerable effort pushing students to make connections between these representations. In terms of their teaching practices for connecting representations, Ms E does this during the whole group discussion before writing stories, and Ms. M does this principally in response to students' stories. Noticeably absent is the use of representations as a purely visual aid designed to make abstract ideas concrete to language learners. In the next section we analyze how these teachers scaffolded the use of language in ways suitable for students from non-dominant linguistic backgrounds, and how this language use helped students interpret the representations as well as the target mathematical concepts.

## Teaching Practices that Support Mathematical Discussion

In the lessons observed, neither teacher used explicit vocabulary instruction, vocabulary drill and practice, or memorization for vocabulary. Instead, both teachers addressed vocabulary in the context of language-rich instruction focused on making sense of the mathematics. Thus, they used many of the research based recommendations for vocabulary development: environments that are language rich, actively involve students in using language, require both receptive and expressive understanding, and require students to use words in multiple ways over extended periods of time (Blachowicz & Fisher, 2000). We also found that both teachers wove together everyday and academic meanings, vocabulary, and ways of talking. Both teachers accepted, built on, and used student contributions to introduce academic language. They often used and built on students' informal contributions and their own informal descriptions to include technical language in their discussions, and used colloquial language to further illustrate their own formal descriptions.

Despite these similarities, the two teachers addressed vocabulary using different approaches. Ms. M began her lesson by clarifying colloquial language used in the problem, for example "traffic tie-ups," and by asking students to describe what they saw on the map (roads, cities, towns, etc). In contrast, Ms. E asked her students to go straight into the simulation without previous discussion of the map or of the meaning of terms used in the problem. Ms. E also made heavy use of gesturing and pointing to the workbook (which was projected on a screen). She used her finger to point to the problem text as she read out loud, and frequently pointed to segments of the graph as the class discussed the problem. Below we provide two examples of the ways that each teacher's practices reflected recommendations for supporting mathematical discussions and the development of academic discourse.

### Example 3: Ms. E (see Figure 3)

This excerpt follows immediately where Example 1 ends. After asking the students to determine "how fast" the bus was going, the teacher used student responses to introduce the word "speed," using the phrase "constant speed" as represented in the graph:

T: How fast was the bus going? Ok we're going hours and miles. Miles and hours.
S: 60 miles per hour.
T: Ok. How did you determine that?
S: Divided the 180 miles by 3 hours.
T: Ok. If I divide the final distance divided by the final time. Ok. That's gonna give me a speed. You're right. But you told me first he was going fast and then he slowed down. So was this a constant speed?
Ss: No.
T: No. Ok. So I need to determine the speed to see how fast he was going before he started slowing down. Ok?

In this example, the teacher moved from a general description of "the bus was going faster then slowed down" (shown in Excerpt 1) to asking students to find the specific speeds at which the bus was traveling. She used the students' responses to her questions to introduce into the discussion mathematical concepts and phrases such as speed, constant speed, and changing speed, all grounded in a discussion of how these three were represented in the graph.

### Example 4: Ms. M (see Figure 4)

The following segment takes place towards the end of the lesson, after students have written stories about two motion situations. This example illustrates how Ms. M moved back and forth between formal and informal descriptions of a horizontal segment and clarified the mathematical significance of that segment on the graph:

T: Now look at the bus. The bus looks pretty normal this time, right? Because we know that right here is a straight, horizontal line right. And what does the flat line tell us?
S: It's going nowhere.
T: It's going nowhere. And what kind of slope is it?
S: Zero.

In this example, Ms. M combined formal academic language to describe the line as "horizontal" with a more informal description of the line as a "flat line." She repeated the student's contribution and followed it with a question about the slope, introducing the formal mathematical concept of a zero slope. Ms. M used colloquial language to lead the class to academic language, interweaving both types of language while grounding the discussion in the graph. In sum, neither teacher used explicit instruction of technical terms, but rather wove together both everyday and academic language to support a discussion about mathematical concepts. Our continuing analysis shows that both teachers' discourse was oriented to mathematical concepts rather than to vocabulary definitions.

## Conclusions

This exploration into the practices of two teachers shows some connections between instruction that is both representation and language rich, and thus supports both conceptual mathematics learning and discussions that are oriented to mathematical concepts. The analysis shows that these two teachers shared some practices but also had different styles, approaches to connecting representations, and ways of addressing vocabulary. We find it promising that this data shows that teachers can have an impact on the mathematics learning of students mistakenly labeled "at-risk." This beginning analysis is also promising in that these teachers are enacting many of the research-based teaching practices for providing representation- and language-rich learning environments.

## Endnotes

(1) We use the term "Latino/a" to be consistent with current usage, although "Hispanic" is the term commonly used in Texas to designate people of Latin American—specifically Mexican—descent.

## References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183-198,

August, D. and Shanahan, T., eds. 2006. *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth.* Mahwah, NJ: Lawrence Erlbaum.

Blachowicz, C., & Fisher, P. (2000). Vocabulary instruction. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (vol. III, pp. 503-523). Mahwah, NJ: Lawrence Erlbaum Associates.

Echevarria, J. Vogt, M. and Short, D. 2004. *Making Content Comprehensible for English Learners: The SIOP Model.* Needham Heights, MA: Allyn & Bacon.

Education Trust. (2003a). *Latino achievement in America.* Retrieved July 14, 2004, from http://www2.edtrust.org/NR/rdonlyres/7DC36C7E-EBBE-43BB-8392-CDC618E1F762/0/LatAchievEnglish.pdf

Gándara, P., & Contreras, F. (2009). *The Latino education crisis: The consequences of failed social policies.* Cambridge, MA: Harvard University Press.

Garcia, E., & Gonzalez, R. (1995). Issues in systemic reform for culturally and linguistically diverse students. *Teachers College Record, 96* (3), 418-431.

Hiebert, J., & Grouws, D. (2007). The effects of classroom mathematics teaching on students' learning. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Reston, VA: NCTM.

Kaput, J. (1994). Democratizing access to calculus: New routes using old roots. In A. Schoenfeld, (Ed.), *Mathematical thinking and problem solving*. Hillsdale, NJ: Erlbaum.

Moschkovich, J. N. (2002). A situated and sociocultural perspective on bilingual mathematics learners. *Mathematical Thinking and Learning, 4*(2 & 3), 189-212.

Moschkovich, J. N. (2007a). Examining mathematical discourse practices. *For the Learning of Mathematics, 27*(1), 24–30

Moschkovich, J. (2007b). Using two languages while learning mathematics, *Educational Studies in Mathematics*, 64(2), 121-144.

Moschkovich, J. N. (2008). "I went by twos, he went by one:" Multiple interpretations of inscriptions as resources for mathematical discussions. *The Journal of the Learning Sciences*, 17(4), 551-587.

O'Halloran, K. L. (2000). Classroom discourse in mathematics: A multisemiotic analysis. *Linguistics and Education, 10*(3), 359-388.

Pierson, J. (2008). *The relationship between teacher follow-up moves and mathematics learning.* Austin: University of Texas. Unpublished doctoral dissertation.

Roschelle, J., Kaput, J., & Stroup, W. (2000). SimCalc: Accelerating students' engagement with the mathematics of change. In M. J. Jacobson & R. B. Kozma (Eds.), *Innovations in science and mathematics education: Advanced designs for technologies of learning* (pp. 47-75). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J., & Gallagher, L. (in press). Scaling up SimCalc: Results from random-assignment field trials in seventh and eighth grade mathematics. *American Educational Research Journal.*

Vahey, P. Lara-Meloy, T., & Knudsen, J. (2009). Meeting the Needs of Diverse Student Populations: Findings from the Scaling Up SimCalc Project. In Swars, S. L., Stinson, D. W., & Lemons-Smith, S. (Eds.). *Proceedings of the 31st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 416-424). Atlanta, GA: Georgia State University.

## Acknowledgments

This material is based on work supported by the National Science Foundation under Grant No. 0437861 and a grant to the Center for the Mathematics Education of Latinos/as (No. ESI-0424983). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

# Interpreting Elementary Science Teacher Responsiveness Through Epistemological Framing

April Cordero Maskiewicz, Point Loma Nazarene University, 3900 Lomaland Dr., San Diego, CA 92106
aprilmaskiewicz@pointloma.edu
Victoria Winters, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182
vwinters@rohan.sdsu.edu

In this study we build on the closely related constructs of teacher attention and responsiveness to explore how one fifth-grade teacher facilitates scientific inquiry. We illustrate the dual use of responsiveness and framing through the case study of Mrs. Charles, who skillfully elicits and builds on students' ideas in her science classroom. In our analysis, we found that ranking the extent of this teacher's responsiveness was inadequate for describing the nature of her teaching and for uncovering ways in which this teacher might progress. Analyzing Mrs. Charles's practice in terms of her framing of the situation, however, began to reveal patterns and nuances in the nature of her responsiveness. We argue that it is through the overlay of her epistemological framing onto an analysis of her responsiveness that we get a fuller picture of Mrs. Charles's practice: inviting and valuing students' contributions while encouraging exploration of certain ideas over others.

## Introduction

Reform in science education calls for active student engagement in scientific inquiry (NRC, 1996), where children propose, evaluate, investigate, and synthesize ideas to develop causal explanations of phenomena (Duschl Schweingruber & Shouse, 2007). While research has documented that children have abundant nascent resources for reasoning about and making sense of the world around them (diSessa, 1993; Metz, 1995; Tytler & Peterson, 2004), transforming these abilities into expertise in scientific inquiry still requires substantial work, both for the learner and for the teacher. Students must engage extensively and productively in practices like reasoning mechanistically (Russ, Scherr, Hammer & Mikeska, 2008), modeling (Schwarz, Reiser, Davis, Kenyon, Acher, Fortus, Shwartz, Hug & Krajcik, 2009), and seeking coherence among ideas (Sikorski, Winters & Hammer, 2009). If students' own ideas and reasoning are to serve as the building blocks for both the process of scientific inquiry and the scientific understandings that constitute the products of inquiry, instruction cannot be scripted or prepackaged (Duschl et al., 2007). Instead, the teacher has the responsibility to elicit, interpret, and follow up on students' reasoning in the moment, in a fashion that values students' ideas as objects of inquiry (Cohen, 2004).

The teacher's role in facilitating scientific inquiry is quite different from what her role would be in a traditional science classroom. Listening and attending to the sense students are making, rather than focusing primarily on how their responses align with scientific canon, requires not only a shift in how the teacher conceptualizes her role in the classroom (Empson and Jacobs, 2008), but contrasts with the social and institutional objectives to which teachers are held accountable (Levin, 2008). Additionally, children frequently have difficulty articulating their thinking, and therefore it requires a teacher's focused attention and effort to interpret student meaning (Sherin & van Es, 2009). The closely related constructs of noticing and responsiveness have emerged in the mathematics education literature as a means of describing what it is teachers attend to in students' thinking and what it is they pursue in their follow-up moves (Empson & Jacobs, 2008; Franke, Carpenter, Levi, & Fennema, 2001; Pierson, 2008; Sherin & van Es, 2009). This literature analyzes teachers' actions as they progress to higher levels of noticing and attending to children's mathematical ideas, emphasizing *what* the teacher responds to without an intentional focus on *why* a teacher might be attending to certain ideas over others. Understanding why a teacher notices or responds to specific ideas or in specific ways can inform efforts to promote teacher change.

In this preliminary study, we explore how to account for the 'what' of a teacher's responsiveness in terms of a possible 'why'—the teacher's framing of the situation. We exemplify how a teacher's facilitation of scientific inquiry can be described in terms of the relationship between her responsiveness and framing, and illustrate this with a case study of one fifth-grade teacher embarking on pedagogical change.

## Framing of the Classroom Activity

Teachers have a tacit understanding of what it means for their students to "do science" that shapes what it is they attempt to enact in their classrooms. A person's answer to the question, "What is it that's going on here?" is what we refer to as his or her *framing* of the situation (Goffman, 1986; Hammer, Elby, Scherr & Redish, 2005; Tannen, 1993). With respect to elementary science, there are many possibilities for how a teacher might answer this question. For example, "doing science" in the classroom may involve reading the textbook, following prescribed experiments, or engaging in debates about natural phenomena. Consistent with Hammer, Russ, Mikeska, and Scherr's (2008) definition of scientific inquiry, we hope teachers will come to see classroom science as the *pursuit of plausible, mechanistic accounts of natural phenomena*.

The theoretical construct of framing, as we use it here, is distinct from but related to a more colloquial use of "framing" as one's intentional portrayal of a situation to others. When we describe how a teacher frames the classroom activity, we are referring to her own perception of the situation rather than to how she deliberately represents a task for her students. For us, the power of framing lies in its ability to describe a person's implicit understanding of the nature of the broader situation she finds herself in. In this sense, we consider framing at a comparatively large grain size. There are multiple aspects to any given framing: social interactions, tool use, and body language, as well as what it means to "know" something in science, that is, the *epistemological* nature of the scientific endeavor. This component of framing differs from epistemological beliefs in that framing accounts for what the teacher does in the classroom, that is, her actions and the roles she takes on while teaching science. This does not always coincide with her explicitly stated, decontextualized views about teaching and the nature of science (Abd-El-Khalick, Bell & Lederman, 1998; Lederman, 1999).

In cases where framing has been used to describe how students reason about science, there is evidence to suggest that their framing constrains both the types of knowledge they employ and how they use that knowledge (Hammer, et al., 2005; Louca, Elby, Hammer & Kagey, 2004). We take this as an indication that a teacher's framing could describe the lens through which she attends to student thinking. This may elucidate how and why she is responsive to certain ideas over others, and how she uses these ideas in her response.

## Responsiveness to Student Thinking

Previous work in mathematics education suggests that how a teacher listens can transform how children talk and, ultimately, influence what they learn (Jacobs, Lamb, Philipp, & Schappelle, 2009; Sherin & van Es, 2009). When studying teachers participating in video-based professional development, researchers found that what teachers attend to—what they notice about the mathematical details of the child's idea—is intertwined with how they respond to students (Empson & Jacobs, 2008; Jacobs, et al., 2009). Classroom studies of teacher noticing use moment-by-moment statements of students and teachers as the units of analysis (Pierson, 2008; Sherin & van Es, 2009). For example, Sherin and van Es (2009) focused on instances when student ideas were presented in class and then determined whether or not teachers noticed these ideas. Pierson (2008) quantified patterns in teachers' follow-up utterances in whole class discussions to characterize a teacher's responsiveness, which she defines as "the extent to which teachers 'take up' students' thinking and focus on student ideas in their moment-to-moment interactions" in the classroom (p.25). Pierson found a significant correlation between student learning and the highest level of teacher responsiveness that she articulates (High II). In defining High II responsiveness, Pierson identifies four types of teacher follow-up moves: an invitation to further explain the idea, offering a contradiction or counterclaim related to the idea, an invitation for other students to make sense of the idea (agree/disagree, etc.), or uptake of the student's idea through revoicing or expanding the reasoning. In our work, we acknowledge the dynamic relationship between attention to students' ideas and responsiveness, and we examine how a teacher responds in order to infer what she is noticing about a student's ideas.

When accounting for the actions of an experienced teacher implementing a module in scientific inquiry, we found that utterances and follow-up questions provided too small of an analytic grain size to adequately make sense of *why* the teacher was choosing to respond in a particular way to students' ideas. Our purpose in this paper, therefore, is to argue that it is more useful to characterize a teacher's progress in promoting scientific inquiry in terms of *both* responsiveness and framing. In our analysis, we aim to show how a teacher's own framing of scientific inquiry influences how she hears and responds to students' ideas.

## Methodology

This study is part of a larger NSF-funded project to, in part, identify what constitutes teachers' progression in their abilities to facilitate scientific inquiry (1). During the first year of professional development, practicing third-through sixth-grade teachers from a large school district in southern California attended a one-week summer workshop in August 2008, and biweekly teacher meetings during the 2008-2009 school year. During these sessions, teachers participated in (a) "science-talks" designed to model what scientific inquiry is, and (b) discussions of classroom video, often taken from their own classrooms, where children's scientific reasoning is on display. We expected that many of the teachers would not initially focus closely on the substance of student reasoning for the reasons described in the introduction. We found, however, that a few teachers did in fact attend to students' ideas in ways that existing measures of responsiveness would categorize as high. For this paper, we focus on one of the latter teachers, Mrs. Charles (pseudonym), and describe how we identified what might constitute progress for her in promoting scientific inquiry in her classroom.

The data sources for this study include video recordings of Mrs. Charles's implementation of a 15-hour module. On the first day of the module, the supporting curricular materials suggest that teachers pose the following question to their students:

> Suppose that one night it rains. When you arrive at school you notice that there are puddles of rainwater in the parking lot. But when you go home you notice that the puddles are gone. What happened to the rainwater?

Rather than follow a scripted curriculum, teachers were encouraged to pursue ideas and questions which students brought up during the discussion of this first question. Additional data included classroom field notes taken by the second author, and video and/or field notes from debriefing sessions with the second author after each day's instruction.

Analysis proceeded in phases. Initially both authors independently viewed all of the classroom video, identifying instances when Mrs. Charles was responsive to students' ideas that were raised in class. We use responsiveness in a broad sense to mean noticing and responding to a student's idea either by rephrasing the idea, probing for further clarification, or shifting the direction of the discussion in a way that addresses the idea (Levin, Hammer, & Coffey, 2009). Finding that Mrs. Charles was generally responsive to students' ideas, each author then identified and transcribed several episodes representing the various types of activities in the classroom (small group discussion, small group experimentation, whole class discussion). Analytic memos summarizing the student ideas and interpreting Mrs. Charles's responses within these episodes were compared by the two authors to try to understand the nature of Mrs. Charles's responsiveness. By means of the cyclical, interpretive analysis cycle (Clement, 2000) (2), we continued to analyze video episodes and modify our account of Mrs. Charles's responsiveness until we found a pattern that fit the data and presented a reasonable, coherent account of her actions. We argue below that it was only by applying the construct of framing to our analysis of Mrs. Charles's actions that we were able to understand why she consistently responded in particular ways to students' ideas. The coupled use of these constructs—responsiveness and framing—is the main focus of this preliminary work, and we save for future analysis the endeavor of quantifying the frequency with which Mrs. Charles frames science in any given way.

## Findings

Mrs. Charles took very quickly to the practice of giving students space to express their ideas. From the initial opening question about a puddle, Mrs. Charles facilitated discussions and encouraged experimentation based on the questions students posed about various topics related to evaporation. For example, after listing on the board several student-generated questions, Mrs. Charles asked: "What do you think that we should explore up there [on the board], that we've been talking about today? And how do you think we should explore it? What do you think we should do?" Students' ideas were welcomed, made public, discussed, and clarified so that they could become objects of inquiry for experimentation. The following segment of transcript, taken from a student-initiated class debate, exemplifies how Mrs. Charles provides space for students to describe their own reasoning and critique others' ideas.

| | |
|---|---|
| Mrs. Charles: | What do you think?! Susan. |
| Susan: | Um, when you were reading it (…) um, the size of the parking lot. I don't think that would affect it, because you're talking about a puddle, which is not, like you're not measuring [pause] |
| Mrs. Charles: | Talk to this table. They were the ones who put it on. I have nothing to do with it. I just read it. [To Cody:] Go. |
| Cody: | Well, um. Well, the heat comes down on the parking lot, 'cause the parking lot is really big, then the heat won't go to that one little water spot. It will go to everywhere, and it won't be hot in that water spot. If it's smaller, it will focus in on that, on a littler spot, and it'll be hotter. |
| ... | [1 min and 40 sec of debate between Cody and Susan] |
| Mrs. Charles: | Andrea, did you want to add something to it? |
| Andrea: | It actually does matter, because, um, if the parking lot's bigger, the sun would go everywhere in the parking lot, and it wouldn't be as hot in that one area where the |

| | |
|---|---|
| | puddle is. And if it's smaller, it would be hotter because it's a smaller area and more heat would be able to go in there, in that one area. |
| Mrs. Charles: | Brian, you're shaking your head. What's going on? |
| Brian: | Wouldn't the sun go to every single place? |
| Susan: | Yeah, but that's what I was doing. |
| Cody: | Yeah, but it wouldn't be as hot, 'cause it has to cover all those places. |

This episode is representative of Mrs. Charles's interactions throughout the module in that she facilitates discussions between students, allowing their reasoning to develop in response to critiques from classmates rather than looking to an authority figure for approval or disapproval. Mrs. Charles promoted experimentation—not the assertion of a teacher or text—as the ultimate authority on "what worked", and often led student discussions toward empirical investigation. During the module, Mrs. Charles guided her class in four cycles of exploration, each consisting of students (a) generating a variety of questions while discussing some aspect of the puddle, (b) selecting which questions to pursue, (c) designing and implementing experiments to test those ideas, (d) reporting outcomes, and (e) finding new questions and topics of investigation. Each cycle lasted approximately 2-4 hours spread over one to two class periods, and students spent a considerable portion of this time engaged in experimenting. This overall organization of the module emerged from Mrs. Charles rather than from professional development or project directives and was driven more by students' ideas than most other teachers' module implementations were during this first year of the project.

Analyzing Mrs. Charles's classes, we find evidence of her responsiveness to student thinking. She consistently responds to students by restating the idea or asking others what they think of the idea without "taking over" the students' thinking. This, according to the literature, is evidence of attention to students' ideas (Empson & Jacobs, 2009; Pierson, 2008; Sherin & van Es, 2009). A closer analysis of extended episodes, however, revealed a consistency underlying Mrs. Charles's responsiveness, a consistency arising from what we suspect to be a relatively stable framing of what science class is all about.

## Framing Influencing Responsiveness

Mrs. Charles's frequent attention to one specific aspect of students' ideas, the potential testability of an idea, is what caused us to question the sufficiency of analyzing Mrs. Charles through the lens of her responsiveness alone. This section provides two extended examples of Mrs. Charles being responsive to her students in a way that promotes science as a fundamentally empirical endeavor. These episodes, although representative of the typical teacher-student interactions in her classroom, were chosen because they effectively portray the nuances of what aspects of student thinking Mrs. Charles notices and responds to, and to what ends she uses these ideas.

### Example 1

While interacting with small groups, Mrs. Charles frequently steers students to consider further experimentation. Just prior to the following episode, Mrs. Charles had asked the whole class to draw what they saw happening to the puddle during their experiments on the previous day. While visiting a small group, she elicits one student's thinking, challenges the epistemological basis of his claim, and then poses a question that both answers that epistemological challenge and suggests a course of action for the students:

| | |
|---|---|
| Mrs. Charles: | Okay, so what is it you have there [on your drawing]? |
| Matt: | Okay. So the heat goes, like, down into the water. And then the water turns into its gashus or gas-e-ous [laughs]--its-its-its gas form--which is called water vapor. And so, like, this is what's happening. |
| Mrs. Charles: | How do you know that it's the heat from the sun that's causing it, and not the heat from the ground? |
| Matt: | Well, it could be the heat from the ground, but the ground got heated up from the sun. So it's originally from-from heat, right? |
| Mrs. Charles: | [to boy sitting next to Matt] So wha'd'ya think? Do you agree with him? |
| Matt: | 'Cause the ground got heat from the sun, and the sun. So it's like both. |
| Mrs. Charles: | How would you test that to prove it? |

Mrs. Charles displays responsiveness throughout this interaction by eliciting ideas, engaging with the specifics of a student's idea, challenging the student's thinking without promoting her own reasoning as a substitute, and inviting others to make sense of that student's idea. Upon closer inspection, we find that Mrs. Charles's responsiveness is focused primarily on the aspects of Matt's idea that lend themselves to empirical investigation.

Matt's response to Mrs. Charles's initial elicitation describes a causal chain of events that begins to account for what happens to the puddle: heat from the sun goes down to the water, which then causes the water to change into its gas form. Mrs. Charles follows up with a responsive, substantive probe ("How do you know … ?") about one specific aspect of Matt's reasoning—that the heat affecting the water comes "down" from the sun. Mrs. Charles's question could be a way of asking Matt to expound on his mechanistic understanding, and perhaps she has reason to believe that the nature of the heat source is vital to his budding model. Her subsequent follow-ups, however, do not continue to press for explanation. Rather, her final comment ("How would you test that to prove it?") is a clear invitation to design an experiment.

The aspect of Matt's reasoning that Mrs. Charles chooses to focus on—where the source of the heat comes from—is perhaps the most practical component of Matt's explanation to investigate empirically. Identifying the source of the heat translates nicely, although not simply, into experimental treatments, where the "ground" surface temperature is controlled, and where a fixed amount of light can reach the puddle. However, a direct empirical investigation of the process Matt alludes to, of how "the water turns into its ... gas form", is not nearly as straightforward. In this sense, Mrs. Charles's highlighting of "heat goes, like, down into the water" rather than "the water turns into its ... gas form" is consistent with a framing of scientific inquiry as fundamentally about developing questions and conducting experiments to answer those questions.

## Example 2

This second example is taken from a group of students in the midst of a controversy: one student is adamant the "stuff" they observed rising off their puddle was heat, and the others insist that it was water vapor. As before, Mrs. Charles elicits students' ideas and follows up on the content of those ideas, and she probes with an epistemological question as opposed to responding to the main point of contention.

| | |
|---|---|
| Mrs. Charles: | K, wha'd'you guys think? |
| Susan: | *She* [Ella] thinks it's water. |
| Ella: | *She* [Susan] thinks it's not water vapor. And then, we think that it *is* water vapor, but then she's gonna raise her hand, say her opinion and stuff- |
| Mrs. Charles: | [to Susan] Why do you think it's not water vapor? |
| Susan: | Because- |
| Mrs. Charles: | What do you think it is? |
| Susan: | Um, I think it's heat because the blacktop is, like, really, like, hot? |
| Mrs. Charles: | Mmm hmm |
| Susan: | And if you put a puddle on it, it's, like, gonna, like, react? to it? |
| Mrs. Charles: | Okay. |
| Susan: | And so, I think what's coming up is, like, heat. |
| Mrs. Charles: | How do you know that if you put a puddle on something hot it's gonna react to it? |

Again, Mrs. Charles displays responsiveness to student thinking in that she elicits student ideas, acknowledges ideas, highlights specific elements of an idea, and challenges the student to support her claim. Even when Ella is reflecting on the nature and ramifications of the group's disagreement, Mrs. Charles chooses to be responsive to the scientific substance of the student's thinking, asking Susan what she thinks the "stuff" is. What is interesting is that Mrs. Charles follows up by probing how Susan would know something would react, as opposed to choosing to pursue the idea under debate, that is, whether the "stuff" is heat or water vapor.

As Mrs. Charles leaves this group, she reinforces that she wants the students to devise an experiment. While the group continues to debate, one member directly asks Mrs. Charles what that "stuff" rising off the puddle is. She responds by saying, "I don't know. What d'you think, and how can we prove what we're thinking?" Here the topic of the implied experiment is central to the controversy: what that "stuff" is. Mrs. Charles asks the student what he thinks the "stuff" is, but instead of asking *why* he thinks that, she focuses on how he could prove what the "stuff" is. Nominating this idea for experimentation provides further evidence of this nuanced aspect of her responsiveness.

## Framing as a way to understand a pattern in Mrs. Charles's Responsiveness

How do we account for the nature of Mrs. Charles' responsiveness and consider what progress for her might look like? In the analyses above, we exemplify how Mrs. Charles seems to attend to one specific aspect of students' ideas—their potential for empirical exploration. We suggest that, in general, interpreting a teacher's actions in terms of her framing of science can elucidate *what* she is responsive to and *how* she is responsive. In the case of Mrs. Charles, this analysis finds her responsiveness taking the form of guiding students towards experimentation as a way to "prove" their thinking. The component of Mrs. Charles's framing that is most relevant here has to do with her stance on how students go about constructing scientific knowledge, which we will refer to as her *epistemological* framing (Hammer, et al., 2005). Along with the consistent nature of her responsiveness to ideas that are testable, additional sources of data (comments about her class during debriefing sessions and the iterative exploration cycles during the module) provide further evidence to suggest that Mrs. Charles frames the inquiry module as an opportunity to pose critical questions about phenomena and investigate them empirically.

The epistemological component of Mrs. Charles's framing treats knowledge as inherently empirical: we have ideas in science, but we have to test them to know if they are correct. This captures an important component of the nature of science: that science is accountable to empirical findings (Abd-El-Khalick & Lederman, 2000; McComas, 1998). Experimentation, however, is only one part of inquiry. Many teachers' conceptualizations of inquiry involve setting up experiments and identifying the observable outcomes, but do not involve the development of theories or models of natural phenomena (Chinn & Malhotra, 2002; Windschitl, 2004). Therefore, teachers' ideas of what science *is* may require subtle, but important changes.

Although the students' ideas and questions provide the terrain for exploration in Mrs. Charles's class, a desirable and sophisticated inquiry practice, we found that her consistent epistemological framing of science inquiry as primarily empirical investigation also resulted in less desirable aspects of classroom science. For example, Mrs. Charles's endorsement of pseudo experiments, such as one group designing a diorama inside a condensation chamber, or another group following online directions to create a cloud in a bottle. Additionally, because the students were encouraged to explore any questions they were interested in pursuing, this resulted in the divergence of different groups of students into different lines of research. Divergent explorations did not ultimately feed into the larger purpose of constructing an explanatory account of evaporation and related phenomena. Therefore, our findings reveal that Mrs. Charles's epistemological framing led her to focus on particular aspects of students' ideas—the testability or potential fruitfulness of the ideas—which resulted in iterative cycles of questions, experimentation, observation, new questions over the course of the 15-hour module. The construct of framing allows us to make sense of *why* Mrs. Charles attended to particular aspects of students' ideas, as well as illuminates a direction for progress for a highly responsive teacher that already views students "as 'havers' of ideas and classroom communities as groups that productively entertain these ideas" (Cohen, 2004, p.21).

## Discussion

The teacher education literature suggests that really listening to and engaging in students' ideas is a characteristic or skill that takes time for teachers to cultivate (Empson & Jacobs, 2008; Franke, Carpenter, Levi, & Fennema, 2001; Pierson, 2008). In an informal interview with Mrs. Charles, she readily admitted that prior to participation in this professional development project she did not enjoy teaching science, and she described her previous science class as "more traditional" because she followed scripted lessons. During her first year of participation in professional development designed to cultivate more sophisticated practices for facilitating scientific inquiry, Mrs. Charles created a science classroom where the students' ideas took center stage. The students' own reasoning provided a basis for developing and pursuing questions surrounding the 'disappearing' puddle.

When analyzing Mrs. Charles's practice in the classroom, we find that she was highly responsive to students' ideas (Pierson, 2008). Mrs. Charles's classroom interactions and follow-ups demonstrated a variety of clearly responsive moves: she provided space for students to articulate their ideas; she took up student thinking by rephrasing, challenging and building on it; she probed for further clarification; and she shifted the direction of the discussion in ways that addressed student ideas. Simply describing the extent to which Mrs. Charles was responsive, however, does not offer a thorough account of Mrs. Charles's facilitation of scientific inquiry. A closer analysis of her responsiveness revealed a pattern: she tended to respond to ideas that lent themselves to empirical testing and often guided students toward experimentation by challenging how they might "know", "prove", or "test" their ideas. Framing helps make sense of the nuances in Mrs. Charles's responsiveness because it can account for *why* Mrs. Charles chose to take up particular ideas and respond in particular ways. In the case of Mrs. Charles, one component of her epistemological framing—her implicit understanding of how students should address the question, "How do we *know* this proposed idea is valid?" while engaging in scientific inquiry—may have contributed to a very specific kind of responsiveness.

This analysis of Mrs. Charles's practice helps us think about what progress for her might entail. Although Mrs. Charles frequently offered glowing testimonials about what her students were doing as a result of her pedagogical changes, Mrs. Charles expressed a concern during a debriefing session near the end of the module implementation. Prior to class she had created posters summarizing the students' questions and explorations throughout the module; however, she had trouble recording what they had *accomplished* or where they had "*arrived*". Considering this dissatisfaction in terms of Mrs. Charles's framing of scientific inquiry provides insight into how to address her concern and also illuminates a potential path for teacher change. While Mrs. Charles skillfully helped students develop and investigate questions, her framing of science as empirical investigation may have limited her ability to recognize opportunities for other components of scientific inquiry. An alternate framing of scientific inquiry as theory construction *and* experimentation may result in Mrs. Charles modifying the ways in which she is responsive. For example, with this expanded framing of inquiry we may see Mrs. Charles lead students in synthesizing their observations in pursuit of a theoretical model of evaporation. While Mrs. Charles encouraged each group to share out the results of their experiments, these findings were not woven back into a collective explanation for evaporation and the water cycle. Instead, the students' empirical findings were sufficient to suggest whether or not a given variable affects the process of evaporation and became the endpoint of that exploration. Mrs. Charles's comments suggest that *she* sees this endpoint as partially inadequate and unsatisfying. Indeed, there is great potential in what Mrs. Charles has already been able to establish in her classroom. Empirical investigation *is* an appropriate way for students to address questions and resolve competing ideas in science. Even the organization of Mrs. Charles's class, where various groups of students pursued different but related topics of investigation, can be a productive model of scientific collaboration and an appropriate means of arriving at scientific knowledge (Brown, Ash, Rutherford, Nakagawa, Gordon & Campione, 1993). By building on Mrs. Charles's existing framing of science, we can help her expand her notion of inquiry to include the coordination of evidence with theory construction, and potentially satisfy her desire to help the students "arrive" at shared explanations for phenomena.

This study is taken from a larger project aimed at, in part, characterizing teacher progress in facilitating scientific inquiry. The construct of *teacher learning progressions* (Schwarz et al., 2009; Thompson, Braaten & Windschitl, 2009), can document how teachers fundamentally alter their view of what can and should take place in the science classroom, and how to hear and develop the beginnings of science in children's thinking. Instead of postulating a progression informed only by data on novice "starting" points and expert "end" points, our larger project attempts to follow the journey of multiple teachers participating in intensive, long-term professional development. Our case study of Mrs. Charles provides a small vignette in that larger story of change. This account of Mrs. Charles's first year can inform how we conceptualize progress and influence what we do during future professional development. For a teacher that easily adopts a student-centered stance in the classroom and is often responsive to student ideas—at least for use in experimentation—defining a learning progression involves more than an analysis of moment-by-moment responsiveness. Understanding Mrs. Charles's epistemological framing of scientific inquiry in the classroom helps illuminate a direction for progress for her, where we can honor and build on what she has already successfully been doing with students.

## Endnotes

(1) The work described in this paper was supported by National Science Foundation grant 0732233, "Learning Progressions for Scientific Inquiry: A Model Implementation in the Context of Energy."
(2) Adapted from Glaser & Strauss' (1967) constant comparison method.

## References

Abd-El-Khalick, F., Bell, R. L. & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education*, *82*(4), 417–436.

Abd-El-Khalick, F. & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: a critical review of the literature. *International Journal of Science Education*, *22*(7), 665–701.

Brown, A. L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A. & Campione, J. C. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 188–228). Cambridge, U. K.: Cambridge University Press.

Chinn, C. A. & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218.

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In R. Lesh & A. Kelly (Eds.), *Handbook of research methodologies for science and mathematics education* (pp. 341-385). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, S. (2004). *Teachers' Professional Development and the Elementary Mathematics Classroom: Bringing Understandings to Light*. Mahwah, NJ: Routledge.

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10*(2 & 3), 105–225.

Duschl, R. A., Schweingruber, H. A. & Shouse, A. W., (Eds). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, D.C.: National Academies Press.
Empson, S. B. & Jacobs, V. R. (2008). Learning to listen to children's mathematics. In D. Tirosh and T. Wood (Eds.), *Tools and processes in mathematics teacher education* (pp. 257–281). Rotterdam, NL: Sense Publishers.
Franke, M. L., Carpenter, T. P., Levi, L. & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal*, *38*(3), 653–689.
Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing Company.
Goffman, E. (1986). *Frame Analysis: An essay on the organization of experience*. New York: Northeastern University Press.
Hammer, D., Elby, A., Scherr, R. & Redish, E. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89–119). Greenwich, CT: Information Age Publishing.
Hammer, D., Russ, R., Mikeska, J. & Scherr, R. E. (2008). Identifying inquiry and conceptualizing students' abilities. In R. A. Duschl and R. E. Grandy (Eds.). *Teaching scientific inquiry: Recommendations for research and implementation* (pp. 138–156). Rotterdam, NL: Sense Publishers.
Lederman, N. G. (1999). Teachers' understanding of the nature of science and classroom practice: Factors that facilitate or impede the relationship. *Journal of Research in Science Teaching*, *36*(8), 916–929.
Levin, D. M. (2008). *What secondary science teachers pay attention to in the classroom: Situating teaching in institutional and social systems.* Unpublished doctoral dissertation, University of Maryland at College Park.
Louca, L., Elby, A., Hammer, D. & Kagey, T. (2004). Epistemological resources: Applying a new epistemological framework to science instruction. *Educational Psychologist*, *39*(1), 57–68.
McComas, W. F. (1998). The principal elements of the nature of science: Dispelling the myths. In W. F. McComas (Ed.), *The nature of science in science education: Rationale and strategies* (pp. 53-70). Dordrecht, NL: Kluwer.
Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research*, *65*(2), 93–127.
National Research Council (1996). *National science education standards*. Washington, D.C.: National Academy Press.
Pierson, J. (2008). The relationship between patterns of classroom discourse and mathematics learning. Unpublished doctoral dissertation, The University of Texas at Austin, 2008.
Russ, R., Scherr, R., Hammer, D. & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*(3), 499–525.
Schwarz, C. V., Reiser, B. J., Davis, E., Kenyon, L., Acher, A., Fortus, D., Shwartz, Y., Hug, B. & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632-654.
Sherin, M. G. & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, *60*(1), 20–37.
Sikorski, T., Winters, V., and Hammer, D. (2009, June). *Defining learning progressions for scientific inquiry*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA.
Tannen, D. (1993). *Framing in discourse*. New York: Oxford University Press.
Thompson, J., Braaten, M. & Windschitl, M. (2009, June). *Learning progressions as vision tools for advancing teachers' pedagogical performance*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA.
Tytler, R. & Peterson, S. (2004). From try it and see to strategic exploration: Characterizing young children's scientific reasoning. *Journal of Research in Science Teaching*, *41*(1), 94–118.
Windschitl, M. (2004). Folk theories of "inquiry": How preservice teachers reproduce the discourse and practices of an atheoretical scientific method. *Journal of Research in Science Teaching*, *41*(5), 481–512.

## Acknowledgements
The authors would like to thank the researchers of the *Learning Progressions for Scientific Inquiry Project* for their suggestions and feedback on this paper.

# Representational Practices in the Activity of Student-Generated Representations (SGR) for Promoting Conceptual Understanding

Orit Parnafes, School of Education, Tel-Aviv University, Israel, oritpa@post.tau.ac.il

**Abstract:** This research aims to investigate student-generation and elaboration of visual representations as a tool for promoting understanding of difficult conceptual domains. The paper focuses on students' naturally occurring representational practices as identified in an activity of student-generated representations (SGR). The research is based on observations of pairs of students, ages 10-14, generating representations while trying to understand the phenomenon of the moon phases. The activity involves a few stages. First, each student generates a representation to promote his or her own understanding of the phenomenon. Then, the students negotiate and co-construct representations with their peers. Finally, they design representations for explaining the phenomenon to an external audience. The analysis identifies various representational practices utilized by students for making sense of the phenomenon, developing explanations, and communicating their ideas to their peers. The analysis examines how these practices support students in achieving some cognitive and communicative goals.

## Introduction

This research aims to investigate student-generation and elaboration of visual representations, as a tool for promoting understanding of difficult conceptual domains. In particular, it examines students' practices of using drawings and visual representations to advance their own understanding.

Generally speaking, the idea of SGR can be motivated by combining two important components that have been shown valuable for learning: First, a wide range of research has shown that *learning with visual representations* enhance learning and understanding (e.g., Larkin & Simon, 1987; Scaife & Rogers, 1996; Ainsworth, 1999; Parnafes, 2007). Second, research has shown that *self-generated explanations* (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, and La Vancher, 1994) promote deeper understanding. This research proposes to combine the two components and investigate the potential of *student-generated visual representations* as a means for explaining difficult phenomena to promote learning and genuine understanding.

Student-generated explanations can take different forms (or representations). In most studies the self-generated explanations are either verbal or textual. Chi et al. (1994) mention that other forms of nonverbal constructive activity, such as diagram drawing, may also be effective at enhancing learning. This is particularly important given that our world is rich in visual images, and this era is characterized in an ever increasing amount of innovative and sophisticated diagrams and visualizations that enhance various types of information in many fields (e.g., Tufte, 2001). It is only sensible to enrich students' repertoire of self-explaining tools to include visual means to enhance their own understanding and learning of researched phenomena. Furthermore, humans think occasionally with images and visual models and those could be expressed in order to be operated on and communicated with. Yet, if we look at school practices, students' opportunities for expression involve, for the most part, verbal and textual forms. Visual forms are mainly offered to students as resources, and rarely as forms for expression and self-generation.

This rational is stimulated also by examining representational practices in scientific areas. Scientists use representations in their practice to promote their own understanding, to think with in order to make scientific progress, and to communicate with other scientists (Latour, 1986; Lynch and Woolgar, 1990; Ochs, Jacoby, & Gonzales, 1994; Nersessian, 2002). DiSessa et al. (1991) studied students' competencies in various representational practices. They show that students, as young as elementary school students, have sophisticated competencies for creating, critiquing and inventing new representations (meta-representational competencies). Developing these competencies, they conjecture, is important in enhancing students' representational innovation, as well as deepening their understanding of any kind of representation (diSessa, 2004). From a meta-representational point of view, the competency of grappling with a tough conceptual field through the *generation and elaboration* of representations was not explored. This may well be an important competency to develop, given that it is a common practice in scientific work, involving a fair amount of inventiveness and creativity.

Research programs that have already been conducted on SGR (e.g., Ainsworth & Loizou, 2003; Roy & Chi, 2005; Cox & Brna, 1995; Cox, 1999; Gobert & Clement, 1999; Hall, Bailey, & Tillman, 1997) demonstrate the increasing interest of the research community in self-generated representations. While much of this research comes from cognitive science, comparing various cognitive differences between experimental settings, there is a growing research that uses qualitative methods for examining activities of students' SGR in an open ended setting (Bamberger, 2007; diSessa, 2004; Danish & Enyedy, 2007; Enyedy, 2005; and Nemirovsky & Tierney, 2001).

The current research explores the proposal that SGR can be a powerful tool for thinking and for developing understanding of difficult topics. One path of exploration of the current research, and the focus of this paper, is examining students' naturally occurring practices of SGR, and analyzing the means by which these practices shape and facilitate their developing explanations and conceptual understanding. The analysis attempts to address the question: In what ways could naturally occurring SGR practices support the achievement of conceptual and communicative goals in the process of developing and elaborating explanations of difficult phenomena?

## Methods

### Data collection

The research is based on observations of 7 pairs of students (see table 1), generating representations while trying to understand the phenomenon of the phases of the moon. The students were 10-14 years old (4-8th grade). Each pair attended a session of one to two hours.

Table 1: the research subjects.

| | Gender | Grade level |
|---|---|---|
| Rose and Natalie | Two girls | 5th grade |
| Merav and Maya | Two girls | 5th grade |
| Tal and Rotem | Two boys | 5th grade |
| Ran and Gil | Two boys | 4th grade |
| Liron and Itai | Two boys | 5th grade |
| Roni and Tom | Two boys | 6th grade |
| Or and Meital | Two girls | 8th grade |

The scientific domain selected for this study is the *Phases of the moon.* The instructional design of the sessions involves four parts, following a beginning, in which the students receive a brief explanation about the research and the session.

1. *An introductory activity:* a brief interview about what the phases of the moon are, the cycle of the moon phases over a month, and the system of moon, Earth and the sun including relations between rotations and revolutions. The interview usually ends with soliciting students' initial explanations about the causes of the phases of the moon.
2. *Personal representations:* the students are asked to draw some representations (diagrams or sketches) to explain the cause of the phases of the moon.
3. *Collaborative representations:* the students are asked to share their representations with one another and to explain the cause of the phases of the moon to their peer based on the representations. Then, they should negotiate and co-construct a shared representation that they both agree on. This part is usually the longest of all parts and the students go through several drafts as they refine their shared understanding.
4. *Presentable representations:* the students are asked to produce a diagram for people that are not present in the activity. Student design a PowerPoint presentation in this phase. In some cases, this stage is conducted in a separate session, a few weeks or even months after the first session.

During the session, the researcher's role is mostly a participatory observer. In principle, interventions are kept to the minimum, and if made, it is for the purpose of clarifying meanings, or asking challenging questions when the students seem to be satisfied with their state of explanation. The sessions are videotaped and then digitized for further analysis. The representations produced are collected and scanned.

### Data analysis

The methodological orientation of this research encompasses a fine-grain detail qualitative analysis of case studies. The theoretical framework is developed in an approach similar to the grounded theory methodology (Glaser & Strauss, 1967; Strauss & Corbin, 1990). The construction of a theory is done by generating categories from evidence taken from a few focus cases. These categories are then explored in other cases, which may support the categorical concept or suggest modifications to make it more generalized.

To conduct the analysis, the various sources of data are examined: 1. The representational forms that the students produced during the session; 2. The video recordings of the interaction between the students as they work together, including their actions such as pointing to aspects of the representations, gesturing, highlighting, arguing, refining, agreeing, disagreeing and so forth. From this close examination some categories emerge concerning *actions and practices of SGR*. These categories were applied on two of the sessions and refined through several iterations of applications. When the framework stabilized and the categories demonstrated usefulness and insight with regards to the issue at hand, the categories were applied on other case studies. The analysis is carried out using the Transana[i] video analysis software.

## Preliminary findings

The preliminary findings are presented in this paper in the form of a set of categories of actions and practices of SGR produced by the analysis. In addition, a sample analysis of one episode is provided to demonstrate how some of the actions and practices support the achievement of conceptual and communicative goals in the process of developing and elaborating explanations using the generation and elaboration of representations.

### The System of Categories

The system of categories includes two sets of categories. The first set includes categories of observed actions operated on the representations, and the second set includes categories of practices for achieving cognitive and communicative goals. Below is a description of each category with some accompanying examples.

#### Observed Actions Operated on the Representations:

A. *Generating representations* - Students draw representations, either from scratch, or, they continue elaborating a representation that already appears on the paper. Students generate representations with various degrees of innovation and inventiveness. Some of their drawings are customary representations of objects and relations between them (see Figure 1 on the left). Generating representations can also be done by borrowing a representation invention or convention from other resources. Resources can include representational aspects drawn by the peer, or a conventional representation seen in a textbook or elsewhere (See Figure 1 in the middle). Ultimately, a representation could be generated by making a representational innovation, in which students use various common signs (circles, lines, words, numbers, and colors) inventively. In Figure 1 on the right, students use big circles to indicate the field of view from different positions on Earth.

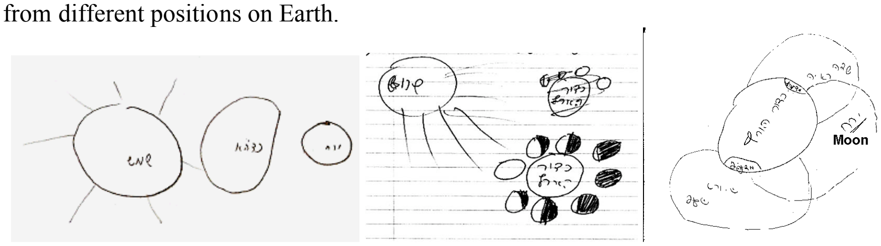



**Figure 1 - various degrees of inventiveness in generating representations**

B. *Gesturing over a representation* - Students make gestures to express various ideas. There are various types of gestures that are found to be used by students through their discussion, including pointing, animating and covering a representation. For example:

a. Animating: Figure 2 shows a selection of animating gestures on the static drawing for representing motion. The two pictures on the left demonstrate gestures that represent the motion of the sun rays from the sun to the moon. The two pictures on the right show gestures that represent the rotation of the Earth around its axis, and the revolution of the moon around the Earth, accordingly.

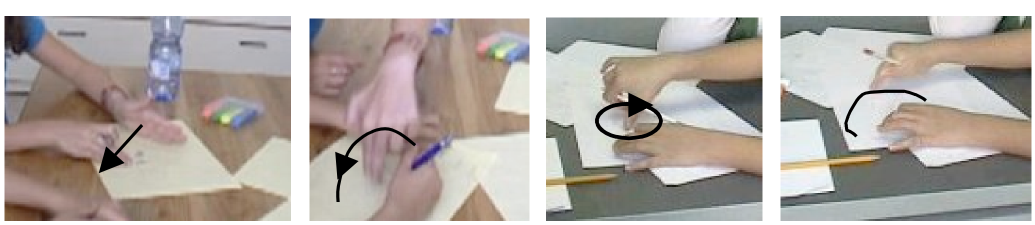


**Figure 2 - animating gestures for representing motion**

b. Covering: Students cover parts of the representations with their hands. In Figure 3 on the left, the student covers parts of the representations to highlight only one moon on which she wants to focus. This is an example of covering and hiding details to reduce destruction from unnecessary details (in this case, drawn by her peer). In the picture on the right, the student covers with her hand half of the moon – the part that is not seen by people looking from the Earth. In this, she tries to reason about the part that is seen from Earth.

**Figure 3 - covering gestures**

C. *Highlighting selected details* - The students select certain details in the representation and highlight them either by circling, shadowing, or making any other form of highlighting. For example, when Natalie talks about the day and night on Earth, she highlights the line splitting the Earth into two halves – the one that's facing the sun, and the one that is not. Another example is highlighting and shadowing a part of the moon that is not facing the sun to show that half of it is shadowed. The students keep highlighting and shadowing the part even though it is already shadowed.

D. *Transforming a representation* - Students transform a piece of an existing drawing by adding details, changing, and deleting. They add details that were not on the original drawing, change some aspects, or delete aspects of the representation. For example, Natalie explains that the moon in the picture cannot be seen from Israel but can be seen from the US. Rose argues with her, saying that even US cannot see it because this moon is dark – the Earth casts shadow on it. She darkens the moon (it was only a blank circle before) and in that transforms the representation to represent a dark moon.

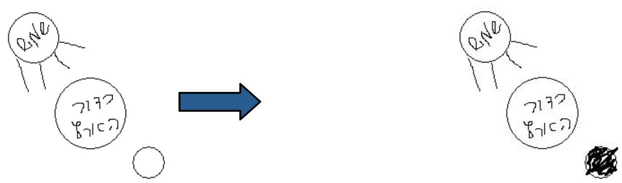


**Figure 4 - Transforming a representation**

As with the "generating representations" category, students transform representations with various degrees of innovation and inventiveness.

## 3.2 Practices Used for Achieving Cognitive and Communicative Goals

Students generate representations, gesture over, highlight and transform representations as a means for achieving some task-related goals. Following are five functions identified in the data for achieving cognitive and communicative goals in this sense making activity.

E. *Organize information* – Students organize on the paper all the relevant information – usually the objects and their spatial relations. This is usually happens at the beginning of producing a drawing.

F. *Construct and communicate an explanation* – Students use representations to generate an explanation either for enhancing their own understanding, or to communicate their explanation to their peer or to the researcher.

G. *Manage complexity* – Students use the representations to reduce or organize epistemological complexity. This can be done by freezing a state in a dynamic process, or by selecting a focus on only one state, by ignoring some details, etc.

H. *"See" better – perceptual aid* - Students use various common signs (circles, lines, words, numbers, and colors) inventively to achieve some communicative goals and to "see" new ideas better.

I. *Claim accountability* - A student uses the representation as evidence for her previous arguments or her peer's previous argument. The representations enable students to make their arguments with reference to the external representations they have produced. They can point to objects in the representations that they wanted to argue about, or to elaborate their ideas on, agree or disagree to statements with reference to them. The instructor can use this practice as well to set the students' attention to a particular issue and to talk about it.

Following is a sample analysis of one short episode. The episode was identified as one in which representations are used for the function of "seeing better" (category H). It is analyzed to reveal how the various actions operated on the representations support the specific cognitive and communicative goal.

## Practices of Using Representations to "see better" – Sample Analysis

Rose and Natalie are 5th grade students. They are towards the end of the third part of the activity, where they made an impressive development in their explanations and understanding. This is their last attempt of explaining the phases of the moon before they move on to the fourth part, where they'd try to generate representations for an external audience. They already explained nicely what happens in the mid month phase (according to the lunar calendar) – when the moon is full. Now, they try to see how other phases are formed. They are now focusing at the location of the moon described in Figure 5:

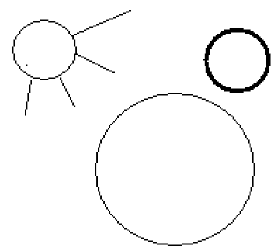
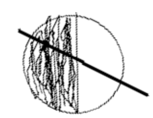


**Figure 5 - The moon in a new location**

Natalie colors the moon in the new location – the part facing the sun is now lit (the colored part is the lighted part). Rose adds a line splitting the moon to two other halves, an important move for highlighting the observer's point of view and how the moon is seen from Earth:

> Rose: So we… what we actually, see, is this part *(drawing a line almost vertical to the colored part)*, more or less…



Her explanation combines both the illumination of the moon (represented by the colored and non-colored parts of the moon), and the part that is seen from a particular position on Earth. She makes a representational innovation, using a line to see the illuminated part seen from Earth. This representational innovation enables the students to "see better" what is actually seen from Earth.

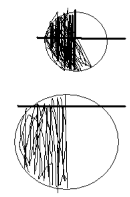
Indeed, following Rose's elaboration of which particular part of the illuminated moon is seen, Natalie gains a meaningful insight:

> Natalie: Ah! One moment, *(takes Rose's hand away from the representation)* the moon doesn't really get larger and smaller, it's simply what we see. We see only the illuminated part [Rose: right!] *(Natalie colors the illuminated part again)*, and now it's like, only half of it, and it's possible that if we'll look before… so maybe it'll be only this part. This line here [Rose: right]. It depends *(looks at me)* on when we look and… when the moon...

This is a moment of "aha" for Natalie. She realizes that the moon does not really get larger and smaller (an idea she kept voicing throughout the session, meaning that only parts of the moon are seen due to occultation), but it is the observed illuminated part seen from Earth that changes. She summarizes by saying that this is a combination of "when we look" (the point of view) and "when the moon" (how it is illuminated). This is a clear case where seeing better leads to a clear development of conceptual understanding.

Rose reiterates the same ideas and combines, more explicitly, the moon illumination and the way an observer sees the moon from Earth:

Rose: So in fact, when the moon is here, so this part is lit *(colors again the half moon that faces the sun)*, and Earth, say, this part *(draws a line on the Earth to show the field of view)* sees relatively this part *(draws a line on the moon to split it – the line is parallel to the line she drew on the Earth)* here, something like that, from the moon, that's lit



Natalie: That's why we see it… These are the phases of the moon!

Rose makes a creative use of signs to emphasize this integration. She colors the half of moon facing the sun, and the other half remains blank to indicate illumination. Next, to represent the observer's point of view, she draws a line splitting the moon into halves, which is parallel to a line representing the field of view of the observer. The half moon facing the Earth is what the observer can see of the moon.

This is an example of how transforming a representation (category D) in an innovative way can facilitate seeing better (category H) of a specific aspect of the phenomenon that is crucial in making sense and understanding it. Rose and Natalie transform the drawing of the moon in a way that highlights and makes visible two important aspects:

1. The illumination of the moon by the sun, represented by the colored part.
2. The part of the moon visible from Earth, represented by the line that splits the moon to the part that faces the Earth, and the part that is hidden from the Earth.

The combination of these two aspects constitutes the core of the explanation of the phases of the moon. The inventive way by which they transformed the representation signifies precisely this important combination.

## Conclusions and further analysis

The analysis classified students' various SGR practices that were used authentically by students when they were asked to generate representations to understand the phases of the moon, and then to communicate their ideas to their peers. These practices support their efforts in thinking, reasoning and making sense of a difficult phenomenon, and maintained their attempt to explain the cause of the phenomenon to their friend, and exchange ideas.

The sample analysis demonstrates the practice of performing certain actions to help one seeing better a particular aspect of the phenomenon. In this case, this was a central aspect at the core of the scientific explanation – the phases of the moon are caused by the combination of the illumination of the moon by the sun and the visibility of the part of the moon from Earth. The students transformed the illustration of the moon in a creative way that expressed this combination, and in fact, helped seeing the shape of the moon seen in the specific phase. The action of "transforming a representation" supports the students' efforts in "seeing better", to facilitate their thinking, reasoning and making sense of a difficult phenomenon, and sustained their attempts to explain the cause of the phenomenon to their friend, and exchange ideas.

In a later analysis, students' trajectories of developing understanding will be described in detail. The SGR practices will then be examined in conjunction to the paths of development, to suggest ways in which such practices could support the process of learning and understanding. Such interrelations will serve to answer questions such as what practices should be fostered, which ones should be discouraged, how specific practices contribute to the development of conceptual understanding? Could some of these practices be taught? How such practices could be supported and enhanced by adequate technological innovations? These can serve as the basis for instructional recommendation of using SGR for science learning.

## Acknowledgments

I wish to thank Andy diSessa, Mariana Levin, David Hammer, Billie Eilam, and Yael Kali for helpful comments related to this work. I thank the participants of the Student-Generated Representation research group at Tel-Aviv University, Orly Shapira, Hadas Shachar, and Meytal Saat, for help in conducting this research. This work was partly supported by a grant to O. Parnafes by the Tel-Aviv university foundation for research promotion.

[i] The data was analyzed with the support of video analysis software – Transana (Wisconsin center for education research, http://www.transana.org). The software was developed specifically for educational research, and it enables uploading video movies, transcribing them, writing notes and interpretations, and creating small clips for easier access and examination.

## References

Ainsworth, S.E & Loizou, A. (2003) The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*, 669-681

Ainsworth, S. (1999). The Functions of Multiple Representations. *Computers and Education, 33* (2-3), 131–52.

Bamberger, J. (2007). Restructuring Conceptual Intuitions Through Invented Notations: From Path-Making to Map-Making. In Teubal (Ed). *The development of notational representations.* New York: Oxford University Press.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P. and Glaser, R. (1989). Self explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2):145-183.

Chi, M. T. H., de Leeuw, N., Chiu, M. H. and La Vancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Cox, R. (1999). Representation construction, externalised cognition and individual differences, *Learning and Instruction 9*, pp. 343–364.

Cox, R. & Brna, P. (1995). Supporting the Use of External Representations in Problem Solving: the Need for Flexible Learning Environments. *The Journal of Artificial Intelligence in Education. 6*(2/3).

Danish, J. A., & Enyedy, N. (2007). Negotiated Representational Mediators: How Young Children Decide What to Include in Their Science Representations. *Science Education, 91*(1), 1-35.

diSessa, A. A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Meta-representational expertise in children. *Journal of Mathematical Behavior,* 10, 117-160.

diSessa, A. A. (2004). Meta-representation: Native competence and targets for instruction. *Cognition and Instruction*, 22(3), 293-331.

Enyedy (2005) Inventing mapping: creating cultural forms to solve collective problems. *Cognition and instruction. 23* (4)

Gobert, J. D., & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching, 36*(1), 39-53.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research.* Chicago: Aldine.

Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words? *Journal of Educational Psychology, 89*(4), 677-681.

Larkin, J., & H. Simon (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–99.

Latour, B. (1986). Visualization and cognition: Thinking with eyes and hands. *Knowledge and Society*, 6, 1-40.

Lynch, M., Woolgar, S., (1990). *Representation in Scientific Practice*. MIT Press: Cambridge, MA.

McNeill, D. (1992). Hand and Mind: what gestures reveal about thought. Chicago: University of Chicago Press.

Nemirovsky, R., and Tierney, C. (2001). Children creating ways to represent changing situations: On the development of homogeneous spaces. *Educational Studies in Mathematics 45* (1-3), 67-102

Nersessian, N. J. (2002). Maxwell and "the Method of Physical Analogy": Model-based reasoning, generic abstraction, and conceptual change. In: *Essays in the History and Philosophy of Science and Mathematics, D. Malament, ed.* 129--166. Lasalle, Il: Open Court.

Ochs, E., Jacoby, S., & Gonzales, P. (1994). Interpretative journeys: How physicists talk and travel through graphic space. *Configurations, 2* (1), pp 151-172.

Parnafes, O. (2007). What does fast mean? Understanding the Physical World through Representations. *The Journal of the Learning Sciences. 16*(3). 415-450.

Roth, W.-M. & McGinn, M.K. (1998). Inscriptions: Toward a Theory of Representing as Social Practice. *Review of educational research 68*(1): 35-59.

Roy, M. & Chi, M.T.H. (2005). The self-explanation principle in multi-media learning. In: R.E. Mayer, Editor, *Cambridge handbook of multimedia learning,* Cambridge University Press, Cambridge, UK, pp. 271–287.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies,* 45, 185–213.

Strauss, A.L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Triona, L. & Klahr, D. (2007) A New Framework for Understanding How Young Children Create External Representations for Puzzles and Problems. In E. Teubal, J. Dockrell., & L. Tolchinsky (Eds.), *Notational Knowledge: Developmental and Historical Perspectives*. Rotterdam: Sense Publishers

Tufte, E. R. (2001) *The Visual Display of Quantitative Information*, Graphics Press, (2nd Ed.).

# Tracing knowledge re-organization - a fine grain analytical framework for looking at students' developing explanations

Orit Parnafes, School of Education, Tel-Aviv University, Israel, oritpa@post.tau.ac.il

**Abstract:** This paper presents an analytical framework for tracing knowledge re-organization through self-generating explanations and representations. The analytical framework, inspired by the knowledge in pieces perspective (diSessa, 1993), is applied on one case study, to illustrate the type of insight the framework could provide. The scientific domain explained by the students in the case study is the phases of the moon. Pairs of students ($5^{th}$ grade/11 year old) met for an hour and a half to generate and elaborate paper and pencil representations in order to explain and develop their understanding of the cause for the phases of the moon. The analysis traces the development of their explanations as an indication of their developing understanding, by tracing the re-organization of knowledge pieces throughout the process.

## Theoretical and methodological approach

This paper presents an analytical framework for tracing knowledge re-organization through the process of self-generating explanations and representations. The analytical framework is based on the knowledge in pieces perspective (diSessa, 1993). The analytical framework is presented using a case study in which two students develop their explanations and representations regarding the phases of the moon. The focus on one motivating case study is used as an opportunity to look into some contentious issues of conceptual change.

Conceptual change is an established research field in which numerous theories exist (e.g., Carey, 1999; Nersessian, 1989; Strike & Posner, 1990; Vosniadou & Brewer, 1992; Wiser, 1988; diSessa, 1993), concerning what students know (specifically about "deep and difficult" topics), how that changes over time, and how to mediate that change. Within the field, there are critical disagreements (diSessa, 2006). One of the critical centers of dispute is the question of the fundamental attributes of the nature of students' intuitive knowledge and how it changes and develops into scientific knowledge.

Some researchers (e.g., Strike & Posner, 1990; Nersessian, 1989; Wiser, 1988) draw parallels between individual's conceptual change and scientific revolutions (Kuhn, 1962). For example, Strike & Posner (1990) argue that students' concepts change from one set of concepts to another, incompatible with the first. Similar to scientific revolutions, what needs to be changed during the process of conceptual change, are some central commitments. Vosniadou & Brewer (1992) maintain that students' knowledge is organized in the form of coherent models and that conceptual change involves a shift from one model to another due to the reinterpretations of the presuppositions that gave rise to the initial models. A considerable part of science and mathematics education research refers to student knowledge in terms of "conceptions" (e.g., misconceptions, preconceptions, or alternative conceptions) that students have about particular concepts. Preconceptions are known to be robust, intact elements of cognitive structure, and demonstrate resistance to instruction. The misconception movement pays little attention to students' productive ideas and the continuities between what students already know and what they need to understand.

The "knowledge in Pieces" theory (diSessa, 1993) suggests a somewhat distinct view on intuitive knowledge and conceptual change. The name "knowledge in pieces" (KiP) suggests an underlying epistemological commitment that knowledge is constituted by a multiplicity of diverse pieces, rather than being a coherent system (e.g., theory-like). Smith, diSessa & Roschelle (1993) provide a constructivist view of learning in which students' ideas play productive rather than destructive roles in the acquisition of expertise, and that the process of conceptual change is a process of reorganization (rather than replacement) in which existing pieces of knowledge are modified in terms of the contexts in which they are activated and used.

The analytical framework presented in this paper, espousing the KiP perspective, provides means for tracing the reorganization of intuitive knowledge elements using a moment-by-moment approach, and demonstrates their productiveness in a more advanced state of understanding. A sample analysis of parts of the case study is presented in order to provide the gist of the way the framework is applied. A detailed analysis is presented in Parnafes (in review).

## The scientific domain – the phases of the moon

Research on students' understanding of the phases of the moon agree that students have difficulties understanding the cause of this phenomenon, and that they usually give a range of alternative explanations (e.g., Sadler, 1987; Baxter, 1989; Barnett & Morran, 2002; Hansen & Barnett, 2004; Kavanagh, Agan & Sneider, 2005; Trundle, Atwood & Christopher, 2007). The literature documents a range of students' alternative conceptions that include the following: The phases of the moon occur due to clouds covering parts of the moon; due to a shadow cast on the moon by the earth; due to a shadow cast on the moon by other planets (Baxter, 1989); or due to a viewer's geographic position on earth, and the supposition that people in different geographic locations see different moon phases (Schoon, 1992; Trundle, Atwood & Christopher, 2007). These alternative conceptions are considered to be fairly robust, hard to change by instruction, and persistent, many times until adulthood (Barnett & Morran, 2002; Baxter, 1989; Trundle, Atwood & Christopher, 2007).

The scientific explanation for the phases of the moon is that half of the moon is illuminated by the sun, and the portion of the illuminated half seen from earth varies as the moon revolves around the earth (Figure 1).

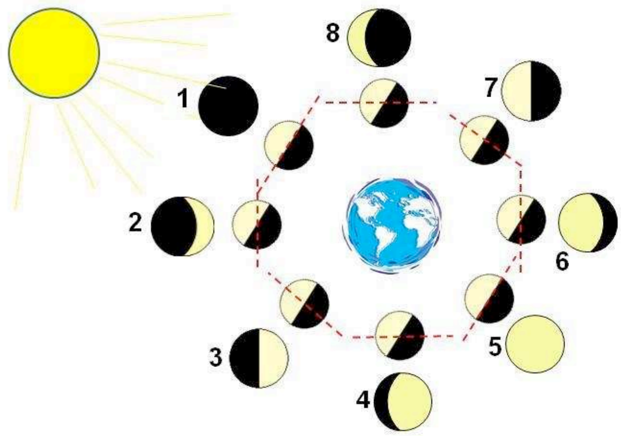



**Figure 1 - The phases of the moon (The internal circle of moons represents the different locations of the moon, illuminated by the sun in the same way throughout the month. The dashed lines show which part of the moon is seen from earth at each location. The external circle of moons represents the way in which the moon is seen from earth in different locations)**

A fair amount of the existing research concerns the classification of students' conceptions about the phases of the moon, and looks at trends of development following specific kinds of instructional interventions, or examines differences among different age groups. A common practice for studying trends of development is developing codes for students' conceptual understanding, and then classifying students' responses based on the type of understanding the codes collectively reflected (e.g., Callison & Wright, 1993; Barnett & Morran, 2002; Trundle, Atwood & Christopher 2007). This paper suggests going beyond this grain level and describing the structure of knowledge and the way it changes in a finer grain level. Taking a finer grain approach means keeping track of more diverse and low level elements rather than a rough unit such as a misconception, and examining how they change on small time-scales. At the fine grain size the "transformation" of a misconception into a scientific conception can be modeled and traced in detail, and in that provide useful instructional implications.

## Methods

### Data Collection

The study presented in this paper is based on observations of pairs of 11 year old (5th grade) students, as they generate representations and explanations and elaborate them while trying to understand the cause for the phases of the moon. During the session, the researcher's role was mostly that of a participatory observer. In principle, interventions were kept to a minimum, and if made, they were made for the purpose of clarifying meanings, or asking challenging questions when the students seemed to be satisfied with their state of explanation. The session was videotaped and then digitized for further analysis, and students' representations were collected and scanned. The sessions lasted an hour and a half and included the following four parts:

1. *An introductory activity:* a brief conversation about what the phases of the moon are, and the model of rotations and revolutions of the earth, moon and sun.
2. *Personal representations:* the students are asked to draw some representations (diagrams or sketches) to explain the cause of the phases of the moon.
3. *Collaborative representations:* the students are asked to share their representations and to explain the cause of the phases of the moon to one another. Then, they co-construct a shared representation.
4. *Presentable representations:* the students are asked to produce a diagram that explains the cause for the phases of the moon to people that are not present in the activity. At this point, the students design a PowerPoint presentation.

### Data analysis and analytical framework

The analytical framework relates students' explanations to their conceptual understanding by considering their explanations as the external performance made on the basis of their conceptual understanding. Their conceptual structure is therefore inferred from the examination of their explanations. The analytical framework describes students' conceptual structure, in the context of developing explanations, as containing (see Figure 2):

1. *An explanatory frame*: a hierarchical structure that provides the rationale for the explanations, consisting of finer grained knowledge pieces such as general schemes and models;
2. *Attended information:* pieces of information that students attend to when attempting to construct explanations. These pieces of information could have a high epistemological status. For example, students rarely argue with direct observations ("I saw the moon several times during the day, so I know it can be seen in daylight",) personal experience ("I know it is night in the US when it is day here, because we make phone calls to my uncle in NY"), an authoritative piece of evidence (facts! "I read it in a science book" or "the teacher told us that…"). Because these pieces of information frequently have a high epistemological status, they support the explanation and contribute to its persistence.

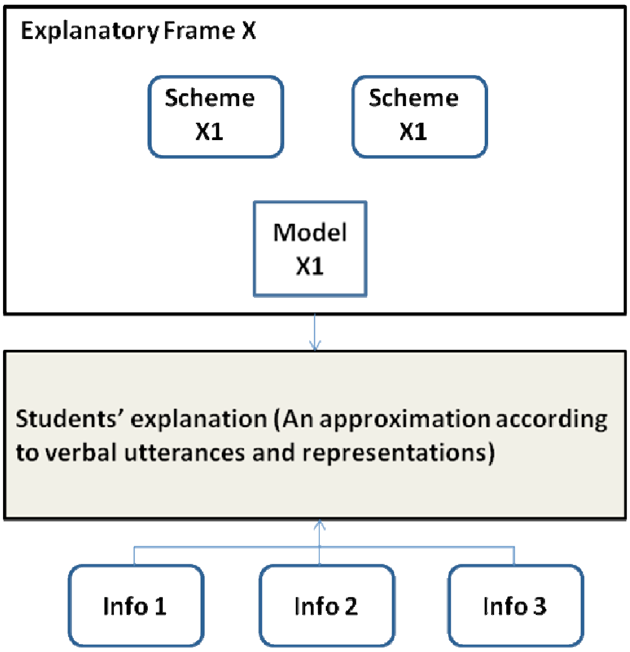



**Figure 2 - A schematic diagram of knowledge pieces used in an explanation**

An explanation is "satisfactory" when the explanatory frame, providing the rationale for the explanation, fits the attended information, or in other words, if these two components feel coherent (Thagard, 1997, Hammer, 2000) for the students at the moment.

A key element in the analysis is the movement from a "satisfactory" explanation to a "better" explanation. A "better" explanation is determined by the quality of coherence, which is characterized by its *resolution* and by its *range* (diSessa, personal communication). Resolution means the amount of available detail used in an explanation. Range means the part of the phenomenology covered by the explanation. In a sense, resolution provides the depth of the explanation and the range provides the breadth. The use of the framework in one case study is demonstrated in the following analysis.

## Analysis and findings

### Rose and Natalie explain the phases of the moon

The students in this study began the session having two different explanatory frames, which they tried to integrate with no success. When the researcher asked a specific question, she apparently provoked a higher resolution, which later on led to a co-construction of a new explanatory frame. A further elaboration of this frame led to a sophisticated explanation that is in line with the normative explanation.

#### Two different explanations for the phases of the moon

At the beginning of the session, after an introductory conversation, each student drew a representation in order to try and personally understand the cause for the phases of the moon (see Figure 3, Figure 4).

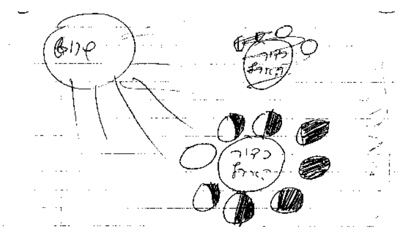


**Figure 3 - Rose's representation**

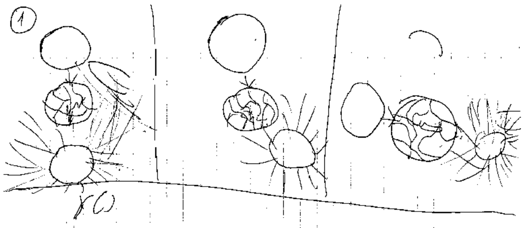
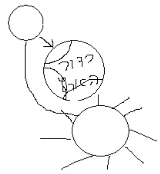


**Figure 4 - Natalie's representation**

When the students were asked to communicate their representations, Natalie focused on an observer's view who sees different parts of the moon when standing in different places on earth:

> …and in the first time, say, the moon is here *(Natalie draws the moon in front of the earth)*, and **only this part of the earth sees the moon** *(Natalie draws an arrow, and an arc on the earth, facing the moon)*, **and this part sees less** *(draws an arc on the side of the earth)*

> This part *(draws another arc on the other side of the earth)*, **this part doesn't see it at all, so in this part it's like there is day** *(Natalie draws the sun)*, **there s sun,**



Natalie explained that in different positions on earth, different parts of the moon can be seen, and integrated into this explanation, the consideration of day and night as experienced on earth.

Rose began generating her representation again, laying out her own explanation as she drew: in the dark phase, the earth blocks the sun rays, hence the moon does not get any light and it is dark:

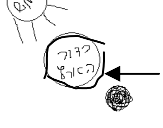
> *(Rose draws the sun and the earth) When the moon is here (draw one moon on the other side of the earth)* **The earth entirely blocks the sun** *(darkens the moon)*, **and then the moon, like, doesn t get any light**

Rose's original representation looks quite like any standard textbook illustration of the phases of the moon: a sun shining on the moon and earth, with multiple moons drawn around the earth. It is likely that Rose was using a convention she'd seen before in a textbook. However, unlike in the textbook illustration, Rose drew the moon closest to the sun (location 1) as a full moon, and the one farthest from the sun (location 5) as a dark moon.

Rose was going to draw other moons around the earth to explain other phases, when Natalie stopped her, saying that Rose's drawing and explanation (so far) were exactly in the same as her explanation. Natalie indicated that, as in her explanation, the moon that Rose had just drawn is the "dark phase." However, while Rose referred to the dark phase as the occasion in which the earth casts a shadow on the moon, Natalie referred to this phase as the occasion in which the moon is not within the observer's field of view.

Natalie So this is like…This is what I said, it's like the day, **we have sun on the entire earth almost,** *(illustrates with her pen the half circle of the earth facing the sun)*, almost, **and the moon, you can't see the moon**

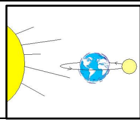
**Except for this part** *(illustrates the half circle facing the moon)*, and there…



Rose: No, it's like, **you cannot see the moon at all, it's dark** *(swiping her hand from the sun to the moon, showing that the earth blocks the sun light)*

**Interpretation:**
The last exchange marks a breakdown in the students' communication based on their two different explanatory frames. The discussion went on for a while, but they each remained within their framework. The following is an analysis of the organization of knowledge pieces that played out in each explanation.

Rose explains the phases within the "casting shadow" explanatory frame that provides the rationale for the explanation. A few schemes are included in this explanatory frame along with a model they operate with:

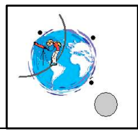
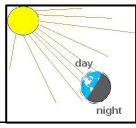
| **Explanatory Frame A** | **"casting shadow"** |
|---|---|
| Scheme A1 | Objects cast shadows in their shape on other objects |
| Scheme A2 | The shadow cast on the obscure object is determined by the alignment of the source of light, the obscure object and the blocking object |
| Model A1 | A moon revolving around the earth, illuminated by the sun  |

The explanatory frame fits with two pieces of attended information, forming a "satisfactory" explanation:

| | |
|---|---|
| Info 1 | The moon looks different every day, varying gradually from an empty to a full moon |
| Info 2 | A textbook illustration (similar to Figure 1) |

The explanation is satisfactory in the sense that it forms low resolution coherence. The explanatory frame fits, to some extent, with the two pieces of information: the earth casts shadow on the moon differently when the moon is in different locations. Yet, Rose does not look at details such as the proportions of phases mapped to percent of shadowing. If she looked into the details of her explanation she would notice that the shadow which is presumably cast on the moon does not correspond to the phases she had drawn, to name one example.

Natalie bases her explanation on two explanatory frames. The first explanatory frame is the "occultation" frame, and the second is the "illumination" frame. A few general schemes construct each explanatory frame along with a model they operate with:

| **Explanatory Frame B** | **"occultation"** |
|---|---|
| Scheme B1 | Partial views of an object are occluded from the observer by obstructing objects. |
| Scheme B2 | What is seen from a particular position is what is within the field of view |
| Scheme B3 | Change in position entails change in the field of view |
| Model B1 | The visibility of the moon from different locations on the earth  |
| **Explanatory Frame C** | **"illumination"** |
| Scheme C1 | A source of light illuminates the part of an object facing it |
| Model C1 | The sun as a source of light illuminating the earth   |

The explanatory frame fits with four attended pieces of information:

| | |
|---|---|
| Info 1 | The moon looks different every day, varying gradually from an empty to a full moon |
| Info 3 | Every day, is constructed of day time and night time |
| Info 4 | When it is day time here, it is night time in countries on the other side of the earth |
| Info 5 | The moon is usually seen at night |

The two explanatory frames fit, to some extent, with the four pieces of information, and in that sense the explanation is satisfactory: different placement positions provide different occultation, and in Natalie's drawing, the moon is apparently seen at night and is obscured during the day. This shows low resolution coherence because Natalie attended to a very general aspect of the phenomenon and tried to account for it, but without getting down to the fine details of the phenomenon. For example, the occultation frame cannot account for the precise shapes of the moon seen throughout the month. In addition, it also shows narrow range coherence - the only one location selected for drawing the moon (location 5) happens to be the location where the moon is seen at night, which fits the pieces of information labeled no. 4 and no. 5. Other locations may challenge the coherence.

At this stage, neither of the students challenged their explanation by getting to a higher level of resolution or extending the range. Each student, presumably, judged her explanation as "satisfactory" in general. When interacting, each student reasoned about her peer's ideas using her own frame.

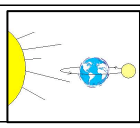
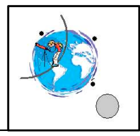
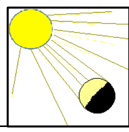
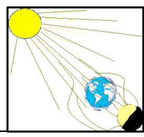
## What happens when the moon is between the sun and the earth?

After a relatively long discussion in which each student explained her representation, the students were asked to produce a shared representation and collaboratively explain the phases of the moon. They made an honest attempt to integrate their explanations; however, they each still reasoned within their explanatory frame and the changes in their explanations were minor. The following instructional move set the stage to a major reorganization. The researcher added another moon to their drawing, between the sun and the earth (location 1), and asked them what is seen at this phase. This particular question was intended to promote further thought that may result in increasing the resolution of their explanations.

OP: **What is seen here?**

Rose immediately said: "When the moon is there, so, like, we see all of it. Right? Because the sun lights it." Her idea that a source of light could actually illuminate an object all around led to a deeper examination of how objects are illuminated exactly, and in particular, how the moon is illuminated in location 1. This discussion resulted in a reorganization of their explanatory frames. The students gradually developed the "combined illumination with point of view" frame, as they discussed the particular details of a specific location of the moon (increasing resolution) and later reasoned about various locations of the moon (increasing range). They kept refining the explanation to the point where a high resolution and a wide range of coherence were achieved, in which point their explanation resembled the scientific explanation. For a detailed analysis see Parnafes (in review).

Rose and Natalie ended the session with an approximately normative explanation of the phases of the moon, saying that half the moon is illuminated by the sun, and the portion of the illuminated half seen from earth varies as the moon revolves around the earth. Although there were still particular ideas that each of them emphasized, they shared this collaborative explanation for the most part. The "combined illumination with point of view" frame is constructed of a few general schemes along with a few models they operate with:

| **Explanatory Frame E** | **"combined illumination with point of view"** |
|---|---|
| Scheme B2 | What is seen from a particular position is what is within the field of view |
| Scheme B3 | Change in position entails change in the field of view |
| Scheme C1 | A source of light illuminates the part of an object facing it |
| Scheme E1 | If the source of light is very big, relatively smaller objects will not cast shadows on other objects |
| Model A1 | A moon revolving around the earth, illuminated by the sun  |
| Model B1 | The visibility of the moon from different locations on the earth.  |
| Model E1 | The sun as a source of light illuminating the moon  |
| Model E2 | A big sun illuminates the earth and the moon  |

Schemes B2, B3 and C1 were used in previous explanations, sometimes in different contexts, with different models. Scheme E1 is used to explain why there is no shadow on the moon in location 5.

The pieces of information attended by both Rose and Natalie, play an important role in the explanation, and they fit with the explanatory frame for the most part. Only one of these pieces of information begins to lose its status. The piece of information labeled no. 5, specifying that the moon is experienced only at night, is acknowledged as inaccurate and as presenting partial information. Natalie emphasized this when she reasoned about the moon in location 5. But as she moved on to think about other locations, she began to comment that the moon is sometimes seen also during the day.

## Discussion

The analysis above was meant to demonstrate the kind of insight the presented analytical framework could bring to issues of conceptual change, and to potential instructional implications. Let us look at some of the issues.

It is still common within the science education community to believe that students' understanding involves alternative conceptions that are robust and are an obstacle for learning the scientific model. The two students in this study might be said to have "misconceptions." In fact, their initial explanations are documented in the literature as common misconceptions: 1. "the phases of the moon occur due to the shadow cast on the moon by the Earth" (e.g., Baxter, 1989); 2. "The different phases of the moon are seen from different parts of the Earth" (e.g., Schoon, 1992).

The above analysis demonstrates a more complicated model of students' understanding, built from finer-grained pieces of knowledge, whereas a configuration of these might appear, roughly, as a misconception. Breaking "misconceptions" into smaller elements, the way that is done in this analysis, suggests productive paths of development.

An important element of the analysis is tracing the knowledge pieces that play out in the explanation. What happens to them when the explanation develops? Are they being transformed? Eliminated? Replaced? It is striking that throughout the session there were not many new schemes used. Most of the schemes in the new explanatory frame E, have already been used in previous explanatory frames B, C and D, and are now re-used in new contexts and with new models. For example, the scheme of illumination (scheme C1), applied previously only on the sun-earth model to reason about the day and night, is now applied on the moon-earth model as well to infer how the moon is illuminated.

The significance of this observation lies in supporting the view that student's intuitive knowledge has continuity and productiveness through the process of conceptual change. It demonstrates that students' explanations are compiled from conceptual resources that are neither right nor wrong, although they may be applied appropriately, inappropriately, in composition with other such schemes or in isolation. The same schemes used in one explanatory frame can be re-configured to form a more adequate explanatory frame. In this sense, knowledge is reorganized (rather than replaced) in such a way that existing pieces are modified in terms of the contexts in which they are activated and used. This also means that students' intuitive pieces are also productive in some contexts and should serve as a basis for constructing scientific knowledge (Smith, diSessa & Roschelle, 1993). In addition to the productiveness of previously used schemes and models, previous explanatory frames may also be productive and useful for explaining other contexts. For example, the casting-shadow frame that is not useful for explaining the phases of the moon, can still be productive for explaining other contexts, e.g., eclipse phenomena. This is a process of shifting contexts, rather than knowledge replacement.

The analysis suggests some instructional implications. If students construct understandings that are composed of a cluster of schemes and models to form an explanatory frame that fits with other elements in their knowledge, they are most likely to respond to a provided "right explanation" by developing two explanations: "an explanation that makes sense" (may look like a "misconception") and "a school explanation" (written in exams with no understanding). The KiP approach suggests students' intuitive ideas should be engaged in order to facilitate continuity of understanding. Increasing resolution and increasing range is one example of a way to support this continuity. Such an approach engages both students' previously constructed knowledge structure, and their intuitive knowledge pieces.

## Acknowledgments

I wish to thank Andy diSessa, Mariana Levin, David Hammer, Billie Eilam, and Yael Kali for helpful comments related to this work. I am grateful for a productive discussion with the Patterns group in UC Berkeley. I thank Orly Shapira for help in conducting this research and for editorial work. This work was partly supported by a grant to O. Parnafes by the Tel-Aviv university foundation for research promotion.

## References

Barnett, M., & Morran, J. (2002). Addressing Children's Alternative Frameworks of the Moon's Phases and Eclipses, *International Journal of Science Education,* 24, 859.

Baxter, J. (1989). Children's understanding of familiar astronomical events, *International Journal of Science Education, 11*, 502.

Callison, P, L. & Wright, E, L. (1993) *The effect of teaching strategies models on Preservice elementary teachers' conceptions about Earth-Sun-Moon Relations.* Paper presented at the Annual meeting of the National Association for Research in Science Teaching, Atlanta.

Carey, S. (1999). Sources of conceptual change. In E. K. Scholnick, K. Nelson, S. A. Gelman & P. Miller (Eds.), *Conceptual development: Piaget's legacy* (pp. 293–326). Mahweh, NJ:

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10* (2–3), 105–225; responses to commentary, 261–280.

diSessa, A. A. (2006). A history of conceptual change research: Threads and fault lines. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 265-281). Cambridge, UK: Cambridge University Press.

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics, Physics Education Research Supplement, 68* (S1), S52-S59

Kavanagh, C., Agan, L., and Sneider, C., 2005, Learning about Phases of the Moon and Eclipses: A Guide for Teachers and Curriculum Developers: *The Astronomy Education Review*, Volume 4, p. 19–52.

Kuhn, T. S. (1962/1996). *The Structure of Scientific Revolutions*. Third Edition. The University of Chicago Press, Chicago.

Parnafes, (in review). Developing understanding as reorganization of knowledge pieces – a case of students explaining the moon phases

Nersessian, N. J. (1989). Conceptual change in science and in science education. *Synthese 80*(1), 163–183.

Sadler, P. M. (1987). Misconceptions in astronomy. In *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics* (pp. 422–425). Ithaca, NY: Cornell University

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3*(2).

Strike, K. A. & Posner, G. J. (1990). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of science, cognitive science, and educational theory and practice*. Albany, NY: SUNY Press.

Thagard, P. (1997) Coherent and creative conceptual combinations. In T.B. Ward, S.M. Smith, & J. Viad (Eds.), *Creative thought: An investigation of conceptual structures and processes*. (pp. 129-141). Washington, D.C.: American Psychological Association. HTML

Trundle, K. C., Atwood, R. K & ,.Christopher, J. E. (2007). Fourth-grade elementary students' conceptions of standards-based lunar concepts. *International Journal of Science Education, 29*(5), 595-616.

Vosniadou, S. & Brewer, W.F. (1994). Mental models of the day/night cycle. *Cognitive Science, 18,* 123-184.

Wiser, M. (1988). The differentiation of heat and temperature: History of science and novice-expert shift. In S. Strauss, (Ed.), *Ontogeny, phylogeny, and historical development* (pp. 28–48). Norwood, NJ: Ablex,

# Perceptions of the relationship between evolutionary theory and biblical explanations of the origins of life and their effects on the learning of evolution among high school students

Pratchayapong Yasri, Rebecca Mancy, University of Glasgow, Glasgow, UK
p.yasri.1@research.gla.ac.uk, r.mancy@educ.gla.ac.uk

**Abstract:** This study explores the perceived relationship between evolution and the biblical account of creation amongst Thai high school students in Christian schools and the impact of these perceptions on learning. Evidence was found for the following perspectives: *creation only* and *compartmental (incompatible models); NOMA*, *fusion*, and *complementary (compatible models)*. Each perspective is related to an approach to learning: *rejection* of one explanation, *switching* between explanations instrumentally according to the context, *differentiating* between areas of reality explained by each, *integration* of the explanations and *refinement* of one explanation through the other. Furthermore, students who held *compatible* perspectives had the potential to engage positively with learning about evolution, those adopting a *rejection* approach tended to engage with evolution in order to falsify it and the student who took a *switching* approach demonstrated detached learning. The study therefore provides an example situation in which learning within one discipline affects learning in another.

## Introduction

Evolution is considered to be a unifying theme in biology; however, its development has relied on the combined effort of professionals in disciplines ranging from biology and medicine to physics and geology, engaged in observing, collection and experimentation to enrich and validate the theory. The area is also becoming ever more interdisciplinary as evolutionary models and frameworks are adopted by those working in domains as diverse as the social sciences and informatics. Research in the area therefore not only produces a body of knowledge that is central to the discipline of biology, it also stimulates scientific collaboration between professionals in different fields; teaching of evolution usually takes place in the context of biology, yet it is central to work in a broad range of discipline areas. Nonetheless, despite its inherent interdisciplinary importance as both a subject of study in its own right and a methodological tool, it has always been the subject of considerable debate, partially fuelled by its relationship with another disciplinary area, that of religion.

Specifically, evolutionary theory is perceived by some to directly contradict religious explanations of the origins of life. There has been a recent renewal of debates pertaining to the teaching of the origins of life and evolutionary theory in formal education (Berkman et al, 2008) with a majority of clashes occurring in relation to Christian groups (Padian 2009). We take the perspective that it is important to understand the ways learners perceive the relationship between religious and evolutionary accounts of the origins of life, as well as their impact on learning, in order to develop teaching approaches which simultaneously respect learners' beliefs and allow them to engage positively with evolution. A range of studies have gathered data on learner beliefs but have largely relied on quantitative approaches using pre-defined categories; we explore the relationships from the perspective of learners, allowing categories to arise from the data.

This paper uses a phenomenographic approach to analyze data collected via individual and in-depth interviews, providing evidence of different understandings of the relationship between evolution and creation as perceived by high school students attending Christian schools in Thailand. The findings reveal five different perspectives of the relationship among the students: *creation only, compartmental, NOMA, fusion* and *complementary* perspectives. Each perspective is related to an approach to learning: *rejection* of one explanation, *switching* between explanations instrumentally according to the context, *differentiating* between areas of reality explained by each, *integration* of the explanations and *refinement* of one explanation through the other. Furthermore, each approach reported here gives rise to different ways of engaging with evolution. Those who adopted *differentiating, integration* and *refining* approaches had the potential to engage positively with the lesson whereas those who adopted *rejection* approach tended to engage the lesson in order to falsify it. Finally, the student who took a *switching* approach demonstrated *detached learning* that appeared to take the form of cognitive understanding with little emotional engagement. A model is proposed that links these findings to perspectives from the literature. More generally, the study suggests that, at least in some contexts, learning of disciplinary material is strongly influenced by learning that takes place within other disciplines, with disciplinary epistemology playing a role in this.

This paper begins by presenting models of the relationship between science and religion, as applied to evolution and creation, drawn from the literature. It then explains the methodological approach implemented in this study, followed by findings, discussion and educational implications.

## Background

The word *evolution* is generally associated with Charles Darwin who described the theory of evolution in *On the Origin of Species* (Scott 2005), and more recent developments of this theory. Evolutionary theory explains changes in species of living organisms over time as due to variation amongst individuals and the processes of natural selection (Darwin 1859). In more recent versions of evolutionary theory, variation is claimed to arise randomly with respect to function, with natural selection acting as a directional force that increases the relative frequency of adaptive traits among the population (e.g. Dawkins 1986).

Another explanation of the origin of life is provided in the Bible and other religious texts. Specifically in the context of Christianity, the Bible explains that God created the Universe, the Earth and all living things (Alexander 2008). The explanation has been interpreted in a variety of ways, ranging from the literal to interpretations of the biblical creation story as a metaphor (Scott 2005). These interpretations have different implications for the relationship with the scientific theory of evolution. Although many Christians see no incompatibility between the biblical account and evolutionary theory (for example, those who see the biblical creation story as metaphorical), some interpretations lead to apparent incompatibilities between the scientific and religious claims regarding origins, with those subscribing solely to the religious viewpoint in its strong form often referred to as "creationists".

Various understandings of the relationship between evolution and biblical accounts of the origins of life have been identified with Alexander (2007) describing four models relating science and religion (*conflict, complementary, fusion* and *NOMA* models). We now provide an overview of these models as a backdrop to our own study. Alexander's work appears to be based on broader philosophical perspectives on the relationship between science and religion (e.g. Barbour 2002), but we concentrate on his work as the most recent source identified.

### 1. The conflict model

The first model identified by Alexander (2007) is the *conflict model*, according to which science and religion are considered to be in fundamental opposition. In this view, the two explanations provide incompatible answers to the same fundamental question. We note that this view can be divided into two: *evolution only* and *creation only* models.

The first subcategory maintains that evolution provides a sufficient and convincing explanation of the origins of life and that there is no place for God or supernatural powers in the scientific arena. This view has been shown to be widespread amongst science teachers and students in a number of settings. For instance, Ingram and Nelson (2006) found that over 60 percent of science students at the Midwestern University accepted the explanation of evolutionary biology, but did not accept biblical interpretations. In contrast, proponents of the *creation only* model reject the theory of evolution and accept that in the Bible (Scott 2005). This model is also found to be accepted by professionals in various domains. For example, Berkman *et al.* (2008) reported that 38 percent of American people stated that they would prefer that creation was taught in schools, not just in addition to, but in place of evolution.

### 2. The "*NOMA*" model

The second model is called *NOMA* after the *Non-Overlapping Magisteria* model proposed by Stephen Jay Gould in 2002 (Alexander 2007). In this view, science and religion deal with different domains of reality. Specifically, Gould (2002) suggests that science and religion focus on different fundamental questions: the magisterium of science relates to the facts of empirical data while the magisterium of religion covers questions concerning ultimate meaning and moral value. The two domains do not overlap by definition thus there is no conflict between the two.

This model corresponds to the viewpoint taken by a range of scientists and has been used by science teachers to support the teaching of evolution. For example, Ladine (2009) notes that separating biblical explanations from the theory of evolution enables him to teach evolution in Christian institutions comfortably and allows Christian students to be open to learning about evolution.

### 3. The complementary model

The *complementary model* provides another way in which evolution and creation can be considered as compatible. Alexander (2007, p.4) explains that "this model maintains that science and religion are addressing the same reality from different perspective, providing explanations that are not in any kind of rivalry to each other, but rather are complementary". For example, one instantiation of this model would be that the belief that God directs the processes of evolution. Proponents of this perspective often argue that putting faith and scientific reason together enables us to better understand the whole range of reality (Berry 2007).

### 4. The fusion model

The final model identified by Alexander (2007) is called the *fusion model*. In this approach, there is no clear distinction between the kind of knowledge provided by evolutionary theory and religious explanations: both focus on the same reality and provide the same knowledge. Alexander (2007, p.3) explains that those who take this perspective "tend to blur the distinction between scientific and religious types of knowledge altogether, or attempt to utilize science in order to construct religious systems of thought, or vice versa".

## Aims of this study

The models identified by Alexander (2007) appear to have been identified from theoretical perspectives from the philosophy of science. Although a number of studies provide evidence for the existence of some of these perspectives, these mainly take the form of evaluations of particular curricula or courses. These studies generally rely on quantitative approaches that offer participants pre-defined categories based on the tutors' own experience or theoretical perspectives drawn from the literature. As a result, it may be that additional perspectives, or variants of these, exist amongst the student population. Furthermore, most empirical studies have been carried out in Western countries (USA, UK, EU countries), thus alternative perspectives may arise in other cultural contexts. Of particular importance to education, the ways in which these perceptions influence learning have not been investigated in detail.

The specific research questions addressed in the study were thus as follows: (1) How do Thai students at Christian schools perceive the relationship between evolution and creation? (2) How does their perception of the relationship relate to their account of the learning process in which they engaged while studying evolution?

The study reported here contributes to the broader literature on the educational implications of perceptions of the relationship between evolution and creation by investigating these viewpoints through discussions with students, without providing pre-defined categories. It focuses on developing a deeper understanding of students' perceptions of the relationship between the accounts of the origins of life, as well as on identifying the impacts of these perceptions on their learning of evolution. Because we expected to encounter less polarized responses in a context where the relationship between creation and evolution has been less politicized than in the Western world, we focus our study on students in Christian schools in Thailand where the Christian population generally adheres to Baptism, Presbyterianism and Pentecostalism. This new cultural context also extends our knowledge to beyond the Western world, potentially allowing additional perspectives to be identified.

## Methodology and method

Since this study aims to identify the different ways that students perceive a specific phenomenon – that of the learning of evolution in their Christian school – phenomenography is chosen as a methodological framework for the study and associated data analysis (Marton 1981). A phenomenographic approach is appropriate for an empirical study such as this that aims to understand the ways in which people experience a particular phenomenon (Marton 1981), with particular emphasis on describing, analyzing and understanding the different ways in which people experience a situation from their personal viewpoint (Micari et al 2007). Researchers using phenomenography normally collect data from individual interviews which are transcribed and analyzed to identify common and contrasting perspectives relating to the phenomena of interest (Marton 1981).

Nine high school students, aged 17-18, currently enrolled in Christian schools in Bangkok took part in this study, alongside 2 teachers. In principle, all students should have encountered evolutionary and biblical explanations of the origins of life in the school context as evolution is a compulsory part of the science curriculum and Christian schools provide Bible Study classes that cover origins. To allow participants the freedom to describe freely the relationship between biblical and evolutionary perspectives on the origins of life, the 30-minute semi-structured interviews employed open-ended questions carried out in their schools. The interview consisted of two parts: firstly, the participants were asked to express their belief concerning the relationship between evolution and creation in a general sense; secondly, they were encouraged to talk about their experiences of learning about these in the classroom.

For data analysis, all interview records were transcribed in Thai. Based on the two sections of the interview, participant narratives were analyzed along three dimensions: their general perceptions of the relationship between the two explanations; the ways in which they approached the task of learning about the two explanations; and the types of their engagement with evolution. Similar and contrasting perspectives relating to each dimension were drawn from the narratives by grouping similar statements into categories to allow common themes to arise, whilst differences were also actively sought between descriptions in order to differentiate between participant perspectives.

Findings were also validated by using *data triangulation* and *investigator triangulation* (Marthison 1988). For data triangulation, we considered both the relationship between our findings and those from the philosophical literature and also interviewed teachers. Teachers and students were interviewed using a similar interview schedule; however, students reported on their own perceptions whilst teachers referred to their own position and to the learning of their students via questions relating to the issues they had encountered when teaching about evolution. For investigator triangulation, two researchers were involved in the data analysis that took place in two phases. The first author identified categories of relationship according to processes of phenomenography. Relevant sections were translated into English and the categories identified were compared to those considered in the literature. In the second phase, the categories were further refined in conjunction with the second author using a process of consensus, until both were comfortable. During the second stage, the categories were compared with Alexander's (2007) distinctions.

## Findings

### 1. The relationship between evolution and creation

Analysis of the student narratives provided five perspectives on the relationship between evolutionary and biblical accounts of the origins of life among participants in this group, summarized in Table 1. Four of the perspectives correspond roughly to categories identified by Alexander (2007), while the other has not been described in this context. The first two categories, *creation only* and *compartmental,* are both *incompatible* perspectives, while the remaining three, *NOMA, fusion* and *complementary* are *compatible* perspectives.

Expressing views in relation to the first of the *incompatible* perspectives, four students (Pavee, Prakhun, Nicha, and Thida) explained that they accepted the biblical explanation of God's creation and rejected evolutionary accounts, considering that it is impossible to reconcile the explanations. Three of these students (Pavee, Prakhun, and Nicha) gave their religious beliefs as the reason for their position, whereas Thida instead emphasized purpose, explaining that belief in God's creation gave her life purpose, whereas evolution did not and she was therefore more comfortable with the former. Their view is named *creation only* and corresponds to the subcategory of Alexander's (2007) *conflict model* where evolution is rejected.

Providing evidence for the second of the *incompatible* perspectives, Praporn explained that she viewed evolutionary and creation perspectives as different, but made no attempt to resolve this apparent contradiction, preferring to use the knowledge relating to each instrumentally in the relevant classes, using the different explanations as appropriate according to the context. This perspective, which we call *compartmental*, is not discussed by Alexander (2007) and we can find no evidence for it in the philosophy of science literature. This is perhaps unsurprising since the philosophy literature focuses explicitly on the relationship between science and religion, whereas the student holding this position actively avoids considering the connection.

In the first of the three *compatible* perspectives, Daungjai and Mothana expressed the view that evolution and creation address different kinds of fundamental question and realities: while evolution deals with the facts of scientific evidence, creation deals with religious concerns, such as belief, values and purpose of life. This perspective is named *NOMA* after Gould's (2002) model. Other *compatible* perspectives were held by Sadudee and Apai who believed that scientific and religious explanations pertain to the same reality, but are nonetheless compatible. Specifically, Sadudee appeared to hold a view most closely aligned with a *fusion* model whereby no clear distinction is made between scientific and religious knowledge, suggesting that evolution can be viewed as scientific creation and creation as scientific belief. Finally, Apai explained that he believed that in time, scientific and religious understandings of the world would become aligned, and that science would confirm his religious beliefs, in a way that is reminiscent of Alexander's (2007) *complementary* perspective.

Table 1: Perceptions of the relationship between evolution and creation

| *Category name* | *Representative statements* |
|---|---|
| Conflict (Creation only) | Since evolution is different from the biblical explanation, I do not believe it. […] Science is too small when compared to God. It cannot prove or disprove the existence of God. (Nicha) |
| | Creation is a purposeful plan of God. I am not purposeless any more in this sense. I know where I am from and where I will go after this. So, I am really happy that I am a beloved daughter of God. Although this idea is like children's thought, it can create my happiness. It is impossible that science can explain everything. (Thida) |
| Compartmental | These two theories seem different in various aspects, like time and process. However, I tried to ignore these differences when I studied. (Praporn) |
| NOMA | To me, evolution is a kind of scientific knowledge, but creation is a kind of religious belief. They are fundamentally different. [Science] deals with the matters and evidence. [Religion] deals with people's ethos and values. It explains what the purpose of life is. It teaches us how to be a good person. (Mothana) |
| Fusion | It is possible that God created small living things and let them evolve until they became billions of species as we can see at this moment. However, God is the origin of life, in my opinion. I do not reject evolution. […] We should open our mind to receive a full range of knowledge. [A Christian biology teacher] can make evolution to be scientific creation and make creation to be scientific belief. (Sadudee) |
| Complementary | I believe in creation because I am a Christian. And also, I believe in evolution because it is reasonable. To me, science is about God and God is about science. […] I treat science as a powerful tool to confirm that there is a God. (Apai) |

## 2. Learning approaches

Five approaches were found to be used to negotiate the relationship when learning about the scientific and religious accounts (see Table 2). Firstly, when Pavee, Prakhun and Nicha (*creation only*), learned about evolution in school, they took the Bible as their starting point, comparing new material to their knowledge of the biblical creation story and rejecting other explanations, in approach we name *rejection*. In addition to declining evolutionary theory, they actively engaged with learning about evolutionary theory with the explicit aim to critique the theory in order to reinforce their own religious perspective. Like these students, although Thida (*creation only*) did not learn about evolution in her formal schooling, she consulted books, websites, and people in the church in order to reinforce her belief and argue against evolutionary theory.

Secondly, Praporn (*compartmental*) referred to having two modes of functioning, turning on “the switch of scientific mode” when learning about evolution, and turning it off again when in religious education classes. In relation to her approach when undergoing assessment, she answered questions about evolution as accurately as possible to gain good marks, despite not believing the explanations provided, or wishing to engage with them in a deep way. We name this approach *switching*.

Thirdly, Daungjai and Mothana (*NOMA*) used what we call a *differentiating* approach, attempting first to differentiate between the realms of applicability of the two accounts, and using the distinction to avoid internal conflict and confusion. Daungjai referred to metaphorical interpretations of the Bible in generating the distinction, whereas Mothana referred to the role of religion being to inform moral decisions and values.

Fourthly, Sadudee (*fusion*) was interested in finding the *integration* of evolution and creation and ways in which to coherently combine evolution and creation, suggesting that perhaps God had created life and that evolution explains the development of life forms into the broad range of species existing today.

Finally, Apai (*complementary*) was interested in learning about evolution and the complexities of life almost as *refinement* of the existence of God, claiming that when aspects of the two perspectives appeared not to correspond, that this was because of the current limitations of our scientific knowledge. For him, science is a powerful tool to allow human beings to see God more clearly.

Table 2: Approaches for negotiating the relationship in learning contexts

| *Approach* | *Representative statements* |
|---|---|
| Rejection | I think it is better to know both explanations. It is because I can know how others think about evolution. I want to know more as I can find reasons to reject it. (Prakhun) |
| | God is the truth. What he says is an ultimate truth. Other explanations that oppose to the biblical account must be wrong. I have to deny them. (Thida) |
| Switching | In a science class, I tried to be a science person by answering all questions about evolution scientifically. When I studied about creation in a religion class, I turned the switch of scientific mode off and turned the religion one on. I tried to be a religious person. I answered everything according to religious beliefs. (Praporn) |
| Differentiating | They have different purposes. So, what I really did was I tried to differentiate both of them first. And then, I tried to understand which aspects that evolution wants to deal with and which aspects that creation tries to imply. (Mothana) |
| Integration | What I am looking for is the intersection between evolution and creation (Sadudee) |
| Refining | My stance is I will not immediately reject what I do not understand. I will think that it might be possible in some way. Although I cannot entirely understand how evolution and creation can be perfectly matched, I think it might be possible in the future. I treat science as a powerful tool to confirm that there is a God. I just think that there is a God and He is the truth. And then, other findings will verify His existence. (Apai) |

## 3. Types of engagement with evolution

Data analysis demonstrated that the perspectives and approaches above can be linked to at least three different types of engagement with evolutionary theory and the understandings that arise from them: *engagement*, *learning to falsify* and *detached learning* (see Table 3).

Firstly, all students who held *compatible* perspectives (*NOMA, fusion and complementary*) appeared to be positively engaged with learning about evolution, claiming that their perspective helped them to avoid confusion and demonstrating enthusiasm and interest in the area. Whilst those with a *NOMA* perspective were interested in evolution principally for the scientific knowledge itself, those taking a *fusion* or *complementary* perspective also saw learning about evolution as a way of clarifying their understandings of religion. Secondly, those who hold a *conflict* view explained that they were actively engaged in their studies or studied outside of the classroom with the explicit aim of learning more about evolution in order to critique evolutionary explanations and thus strengthen arguments for their religious viewpoint, engaging with the primary purpose of *learning to falsify*. Finally, the student who took a *compartmental* view demonstrated *detached learning*, only acquiring knowledge about evolution because it was part of the school curriculum.

Table 3: Types of engagement with evolutionary theory

| *Type* | *Representative statements* |
|---|---|
| Engagement | I did not feel uncomfortable or want to reject it. In contrast, it was exciting to know how evolutionary processes work. The more I can see how science is complicated, the more I am sure that there is a |

| | |
|---|---|
| | God. As the Bible says, God is the source of wisdom. He gave the wisdom to us in order to let us see His works. (Apai) |
| Learning to falsify | There was no negative impact on my study at all. I even could get a good grade in that term. Basically, I am confident with God's creation. But I also wanted to know about evolution so that I can argue against it. (Pavee) |
| | At that time, I did nothing, as I did not want to create any controversial issue in the class. However, in my mind, I already rejected it. Then, I tried to ask my friend's father and other people in the church about this. They gave me a lot of information. […] In addition, I read many books that support what I believe. I tried very hard to protect my belief. I stopped doing that when I felt that I did not wonder about the existence of God anymore. (Nicha) |
| Detached learning | In a science class, I tried to be a science person by answering all questions about evolution scientifically. When I studied about creation in a religion class, I turned the switch of scientific mode off and turned the religion one on. I tried to be a religious person. (Praporn) |

## Discussion and implications

In relation to earlier work, the study described here allows for the identification of additional perspectives not demonstrated in other empirical research. Based on the findings and in relation to the four models proposed by Alexander (2007), we claim that there are two primary ways of viewing the relationship between evolution and creation: *incompatible* and *compatible*. Figure 1 provides an overview of the findings, linking together perspectives on the relationship between evolution and creation, approaches to learning and engagement with evolution, as well as situating these within other perspectives in the literature.

In the *creation only* perspective, creation is accepted and evolution is not considered to provide an acceptable explanation; in the *evolution only* perspective, evolution is judged to be plausible and religious accounts are rejected; in the *compartmental* perspective, both explanations are considered to be internally consistent but are in conflict with one another, with no decision made as to the veracity of each. Although the first two models are discussed by Alexander (2007) and others, the *compartmental* model appears not to have arisen in earlier work and we suggest that it should be added as a subcategory of the *incompatible* perspectives. Furthermore, the data collected in this study therefore demonstrate the existence of incompatible amongst Christian students in Thailand, despite the background of Buddhist communities and the less politicized context.

The findings also reveal at least three categories of *compatible* perspectives, corresponding to those identified by Alexander (2007): *NOMA*, *fusion* and *complementary*. Students taking the *NOMA* perspective consider that evolution and creation are addressing different kinds of question and through seeking the different aspects of reality to which each explanation applies, avoid conflict. Students who hold a *fusion* view, do not seek distinctions, but rather attempt to integrate the two explanations through blending, perceiving no clear distinction between them. Finally, students taking a *complementary* viewpoint look for complementarities between the explanations and consider them to be mutually supportive. The data provide empirical evidence for the existence of all three compatible perspectives amongst Thai high school biology students.

The categories indentified above were triangulated by interview data given by two biology teachers working at the Protestant school who had both been teaching biology for at least 5 years, and at least one of whom had taught evolution to the four students from the sample who were studying at this school. The teachers themselves appeared to hold a *NOMA* perspective. Moreover, they provided a third-person perspective on student learning describing examples of situations that correspond to *conflict*, *fusion* and *complementary* perspectives amongst their students. For example, one teacher stated that 'some of my students declined answering exam questions scientifically but they wrote their point of view instead, for example, God created all things (*conflict* perspective) and that 'some of them answered the questions according to scientific explanations discussed in the class but they concluded that all of those are controlled by God (*fusion or complementary*).

The different perspectives and associated learning approaches identified appear to give rise to a range of ways of engaging with evolutionary theory, which in turn support different educational outcomes for the students (see Figure 1). According to our proposed model, those students taking one of the *compatible* perspectives (*NOMA, fusion* and *complementary*) are able to learn about evolution without any conflict with their religious beliefs. These perspectives support either neutral or positive relationships between the learning that goes on in these two disciplines. However, students holding the *incompatible* viewpoints demonstrate different kinds of influences on their understanding and engagement. Not surprisingly, it was among students who interpreted the Bible literally that serious misconceptions (Scott 2005) were identified:

> Evolution is about monkeys and humans. It claims that we are from monkeys. Natural selection is the reason why those monkeys could be humans. Those monkeys tried to adapt themselves to environments in which they lived. (Thida)

More subtly, despite rejecting evolution, some students who *learn to falsify* (and potentially also those using *switching*) may demonstrate convincing understanding of evolution:

> There was no negative impact on my study at all. I even could get a good grade in that term. Basically, I am confident with God's creation. But I also wanted to know about evolution so that I can argue against it and answer questions related to the topic. […] I actually should feel bored while I was studying. But it is not true. On the contrary, I was really keen to learn it. […] It is weird, I know, that I wanted to know in order to reject it. (Pavee)

Although Pavee appears to have understood evolutionary theory sufficiently to do well on the course, his learning is qualitatively different from that of students who accept evolution as accurate.

In his introduction to the *Cambridge Handbook of the Learning Sciences*, Sawyer (2005, p.2) emphasizes the educational importance of "deeper conceptual understanding", contrasting this with factual and procedural knowledge. However, the comments from students and teachers indicate that conceptual understanding does not necessarily lead to acceptance of evolutionary theory. As educators, we need to ask the question of the kind of learning – conceptual understanding or acceptance – we aspire to engender, as well as the moral implications of these intentions in a context where disciplinary knowledge may potentially threaten belief systems that are important to learners. There is evidence to suggest that some pedagogical approaches can support the acceptance of evolutionary theory without threatening religious beliefs. However, it is interesting to recognize that although accepting evolution as plausible may be a prerequisite to studying evolutionary biology, it may be possible to use evolutionary approaches as a methodology in other areas (such as the social sciences) whilst rejecting biological evolution. Indeed, approaching evolution from this perspective may even provide new pedagogical approaches.

At present, one of the most promising pedagogical approaches, as described in Ladine (2009), is the explicit exploration of the NOMA model with students. Ladine (2009) explains to students that he values religious studies as moral development and science as the discipline for naturalistic explanation, and provides evidence that this approach gives rise to positive engagement with evolution amongst his students. In an open-ended question that formed part of a course evaluation questionnaire, students of a private Christian college in the US explained that they had learnt that "evolution doesn't take God out of the picture," "learning that evolution and religion can coexist," and "that evolution may not really be as much of an evil lie as I have always been taught" (Ladine 2009, p.391).

Alternative pedagogies based on the other approaches associated with *compatible* perspectives – *integration* and *refining* – may also be developed. However, it is necessary tread carefully in this area. It was slightly surprising to us to realize that although three students (Pavee, Nicha and Sadudee) noted that evolutionary theory explains changes in species over time and is therefore not strictly concerned with origins of life (biogenesis), two of these students (Nicha and Sadudee) nonetheless rejected evolution. In other words, being explicit about this particular aspect of integration is not sufficient to allow students to accept evolutionary theory as they can nonetheless see conflicts in relation to other aspects, including timescale and purpose.

Even if we do not see the need for students to appropriate evolutionary theory, using an approach such as that described by Ladine (2009) may nonetheless be easier to justify on ethical grounds than leaving students to generate their own understandings of the relationship, as this can lead to a state of uncomfortable confusion:

> To be honest, I am confused, as I don't know what is true. However, I try to ignore that. I try not to think about it too much. As I know, if I ask biology teachers, of course, they will use scientific evidence to explain it to me. If I ask religious teachers, they are going to let me read what the Bible says. There is no way to find their compatibility. So, it is better to keep quiet and forget it. (Praporn)

Finally, however, it is our hope that explicitly addressing the relationship between the two disciplines in order to encourage more sophisticated understandings of this may be to engender greater respect amongst learners in the classroom.

It is important to note some limitations of this study. The dataset was limited primarily due to political unrest in Thailand in 2009 at the time of data collection, making data collection from a larger sample, or indeed follow-up interviews, impossible. A single interview with each participant, is, however, reasonable in the context of this study since we intended to capture student positions at a single time point, rather than to establish how these changed over time. Since we were interested in their perceptions (a first-person account), the primary data source was student interviews, although two teachers were also interviewed. The nature of phenomenography, which aims primarily to develop well-defined categories of experience of a particular phenomenon, means that at least at early stages, these may constitute only a subset of categories that exist, and that furthermore, individuals included in the sample may be, in fact, unique. This does not detract from the validity of the study itself, but makes it difficult to generalize the results until further investigation has been carried out. Our claim is that the perspectives described have been uncovered; however, we make no claim as to their relative prevalence in the population either amongst the larger population from which they were drawn (high schools students at Christian schools in Thailand) or more generally.

More generally, the findings reported here reveal that at least in some areas, learning within one discipline can be strongly influenced by knowledge and learning within another: specifically, theology is shown to impact on scientific understandings. In line with Gould's (2002) claims about the different types of knowledge associated with religion and science, this impact may have its roots in the epistemology of science and religion more generally.

Indeed, Ladine (2009), while discussing perspectives of evolution and creation asserted that including God in the explanation of science indicates misunderstanding of the nature of science. It is therefore interesting to note a tendency amongst those students holding *compatible* perspectives to refer to science and religion more generally in their comments, despite being asked about evolution and creation.

For researchers, this study demonstrates the importance of an interdisciplinary approach to the study of learning in real world contexts. Specifically, it highlights the significance of disciplinary understandings of the nature of knowledge and reality, and informs theory in the learning sciences by exploring their effects on conceptual understanding of disciplinary material. In order to support student understanding and acceptance of disciplinary knowledge, we need to view our practice not as isolated within a discipline, but take account of its position in relation to knowledge and epistemologies appropriate to other disciplines. As educators, we need to be aware of the possible interactions between the epistemological paradigms underlying the disciplines in which we teach in order to address any conflicts sensitively and intelligently.

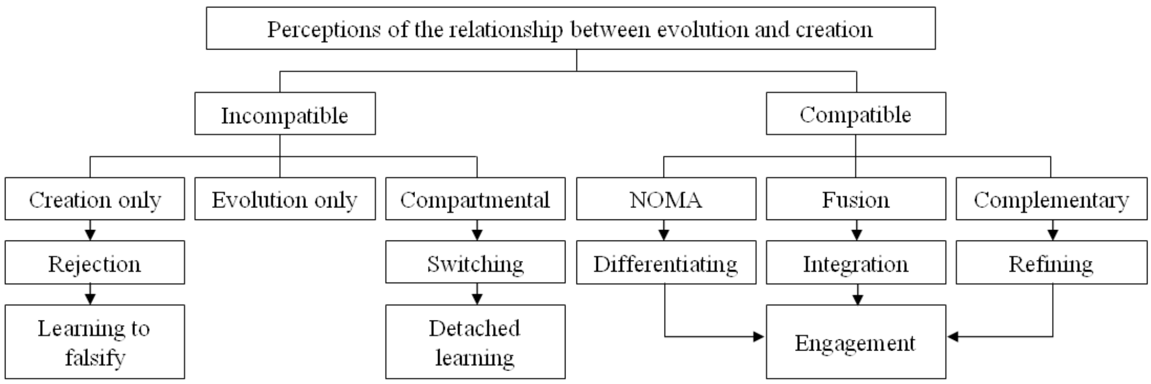



Figure 1: Proposed model relating the relationship between evolution and creation

## Acknowledgment

We would like to thank to the Royal Thai Government for the financial support provided to the first author throughout this study.

## References

Alexander, D. R. (2007). Models for relating science and religion. *Faraday paper*, 3, 1-4.
Alexander, D. R. (2008). *Creation or evolution: Do we have to choose?* Oxford: Monarch Books.
Barbour, I. G. (2002). *Nature, Human Nature, and God*. Norristown, Pennsylvania: Fortress Press.
Berkman, M., Pacheco, J. and Plutzer, E. (2008). Evolution and creation in America's classroom: A national portrait. *PLOS Biology*, 6(5), 920-924.
Berry, R. (2007). Creation and evolution: Not creation or evolution. *Faraday paper*, 12, 1-4.
Darwin, R. C. (1859). *The origin of species by means of natural selection*. London: John Murray.
Dawkins, R. (1986). *The blind watchmaker*. New York: Norton.
Gould, J. (2002). *Rock of ages: Science and religion in the fullness of life*. New York: Ballantine Books.
Ingram, E. and Nelson, C. (2006). Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course. *Journal of Research on Science Teaching*, 43(1), 7-24.
Ladine, T. (2009). Attitudes of students at a private Christian liberal arts university towards the teaching of evolution. *Evolution: Education and Outreach*, 2(3), 386-392.
Marthison, S. (1988). Why triangulation? *Educational Researcher*, 17(2), 13-17.
Marton, F. (1981). Phenomenography: Describing conceptions of the world around us. *Instructional Science*, 10, 177-200.
Padian, K. (2009). The evolution of creationists in the United States: Where are they now and where are they going? *Comptes Rendus Biologies*, 332, 100-109.
Sawyer, R. K. (2006). Introduction: The new science of learning. In R. K. Sawyer (Ed), *The Cambridge handbook of learning sciences* (pp. 1-16). Cambridge: Cambridge University Press.
Scott, E. (2005). *Evolution vs. creationism: An introduction*. California: University of California Press.

# Interactional Achievement of Shared Mathematical Understanding in a Virtual Math Team

Murat Perit Cakir, Gerry Stahl, Alan Zemel
Drexel University, 3141 Chestnut St, Philadelphia PA
mpc48@drexel.edu, gerry.stahl@drexel.edu, arz56@drexel.edu

**Abstract:** Learning mathematics involves specific forms of social practice. In this paper, we describe socially situated, interactional processes involved with collaborative learning of mathematics in a special online collaborative learning environment. Our analysis highlights the methodic ways group members enact the affordances of their situation (a) to visually explore a mathematical pattern, (b) to co-construct shared mathematical artifacts, (c) to make visible the meaning of the construction, (d) to translate between graphical, narrative and symbolic representations and (e) to coordinate their actions across multiple interaction spaces, while they are working on open-ended math problems. In particular, we identify key roles of referential and representational practices in the co-construction of deep mathematical group understanding. The case study illustrates how mathematical understanding is built and shared through the online interaction.

## Introduction

Developing pedagogies and instructional tools to support learning math with understanding is a major goal in mathematics education (NCTM, 2000). A common theme among various characterizations of mathematical understanding in the math education literature involves *constructing relationships* among mathematical facts and procedures (Hiebert & Wearne, 1996). In particular, math education practitioners treat recognition of connections among *multiple realizations* of a math concept encapsulated in various inscriptional forms as evidence of deep understanding of that subject matter (Kaput, 1998; Sfard, 2008; Healy & Hoyles, 1999). For instance, the concept of function in the modern math curriculum is introduced through its graphical, narrative, tabular, and symbolic realizations. Hence, a deep understanding of the function concept is ascribed to a learner to the extent he/she can demonstrate how seemingly different graphical, narrative, and symbolic forms are interrelated as realizations of each other in specific problem-solving circumstances that require the use of functions. On the other hand, students who demonstrate difficulties in realizing such connections are considered to perceive actions associated with distinct forms as isolated sets of skills, and hence are said to have a shallow understanding of the subject matter (Carpenter & Lehrer, 1999).

*Multimodal interaction spaces*—which typically bring together two or more synchronous online communication technologies such as text-chat and a shared graphical workspace—have been widely employed in CSCL research and in commercial collaboration suites such as Elluminate and Wimba to support collaborative learning activities of small groups online (Dillenbourg & Traum, 2006; Soller, 2004; Suthers et al., 2001). The way such systems are designed as a juxtaposition of several technologically independent online communication tools not only brings various *affordances* (i.e. possibilities-for and/or constraints-on actions), but also carries important interactional consequences for the users (Cakir, Zemel & Stahl, 2009; Suthers, 2006; Dohn 2009). Providing access to a rich set of modalities for action allows users to demonstrate their reasoning in multiple semiotic forms. Nevertheless, the achievement of connections that foster the kind of mathematical understanding desired by math educators is conditioned upon team members' success in devising shared methods for coordinated use of these rich resources.

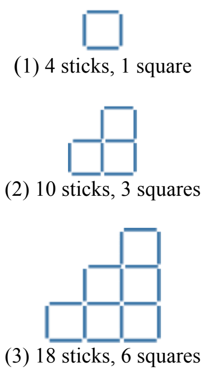
Although CSCL environments with multimodal interaction spaces offer rich possibilities for the creation, manipulation, and sharing of mathematical artifacts online, the interactional organization of mathematical meaning-making activities in such online environments is a relatively unexplored area in CSCL and in math education. In an effort to address this gap, we have designed an online environment with multiple interaction spaces called Virtual Math Teams (VMT), which allows users to exchange textual as well as graphical contributions online (Stahl, 2009). The VMT environment also provides additional resources, such as explicit referencing and special awareness markers, to help users coordinate their actions across multiple spaces. Of special interest to researchers, this environment includes a Replayer tool to replay a chat session as it unfolded in real time and inspect how students organize their joint activity to achieve the kinds of connections indicative of deep understanding of math.

In this paper we focus on the practical methods through which VMT participants achieve the kinds of connections across multiple semiotic modalities that are often taken as indicative of deep mathematical understanding. We take the math education practitioners' account of what constitutes deep learning of math as a starting point, but instead of treating understanding as a mental state of the individual learner that is typically inferred by outcome measures, we argue that deep mathematical understanding can be located in the practices of collective multimodal reasoning displayed by teams of students through the sequential and spatial organization of their actions. In an effort to study the practices of multimodal reasoning online, we employ an ethnomethodological case-study approach and investigate the methods through which small groups of students coordinate their actions across multiple interaction spaces of the VMT environment as they collectively construct, relate and reason with multiple forms of mathematical artifacts to solve an open-ended math problem. Our analysis has identified key roles of referential and representational practices in the co-construction of deep mathematical understanding.

## Data & Methodology

The excerpts we analyze in this paper are obtained from a problem-solving session of a team of three upper-middle-school students who participated in the VMT Spring Fest 2006. This event brought together several teams from the US, Singapore, and Scotland to collaborate on an open-ended math task on combinatorial patterns. Students were recruited anonymously through their teachers. Members of the teams generally did not know each other before the first session. Neither they nor we knew anything about each other (e.g., age or gender) except chat handle and information that may have been communicated during the sessions. Each group participated in four sessions during a two-week period, and each session lasted over an hour. Each session was moderated by a Math Forum staff member; the facilitators' task was to help the teams when they experienced technical difficulties, not to instruct or participate in the problem-solving work. Figure 6 below shows a screenshot of the VMT Chat environment that hosted these online sessions.

During their first session, all the teams were asked to work on a particular pattern of squares made up of sticks (see Figure 1 below). For the remaining three sessions the teams were asked to come up with their own shapes, describe the patterns they observed as mathematical formulae, and share their observations with other teams through a wiki page. This task was chosen because of the possibilities it afforded for many different solution approaches ranging from simple counting procedures to more advanced methods, such as the use of recursive functions and exploring the arithmetic properties of various number sequences. Moreover, the task had both algebraic and geometric aspects, which would potentially allow us to observe how participants put many features of the VMT software system into use. The open-ended nature of the activity stemmed from the need to agree upon a new shape made by sticks. This required groups to engage in a different kind of problem-solving activity as compared to traditional situations where questions are given in advance and there is a single "correct" answer—presumably already known by a teacher. We used a traditional problem to seed the activity and then left it up to each group to decide the kinds of shapes they found interesting and worth exploring further (Moss & Beatty, 2006; Watson & Mason, 2005).




| N | Sticks | Squares |
|---|---|---|
| 1 | 4 | 1 |
| 2 | 10 | 3 |
| 3 | 18 | 6 |
| 4 | ? | ? |
| 5 | ? | ? |
| 6 | ? | ? |
| ... | ... | ... |
| N | ? | ? |

**Session I**

1. Draw the pattern for N=4, N=5, and N=6 in the whiteboard. Discuss as a group: How does the graphic pattern grow?
2. Fill in the cells of the table for sticks and squares in rows N=4, N=5, and N=6. Once you agree on these results, post them on the VMT Wiki
3. Can your group see a pattern of growth for the number of sticks and squares? When you are ready, post your ideas about the pattern of growth on the VMT Wiki.

**Sessions II and III**

1. Discuss the feedback that you received about your previous session.
2. **WHAT IF?** Mathematicians do not just solve other people's problems - they also explore little worlds of patterns that they define and find interesting. Think about other mathematical problems related to the problem with the sticks. For instance, consider other arrangements of squares in addition to the triangle arrangement (diamond, cross, etc.). **What if** instead of squares you use other polygons like triangles, hexagons, etc.? Which polygons work well for building patterns like this? How about 3-D figures, like cubes with edges, sides and cubes? What are the different methods (induction, series, recursion, graphing, tables, etc.) you can use to analyze these different patterns?
3. Go to the VMT Wiki and share the most interesting math problems that your group chose to work on.

**Figure 1:** Task description for Spring Fest 2006

Studying the collective meaning-making practices enacted by the users of CSCL systems requires a close analysis of the process of collaboration itself (Stahl, Koschmann & Suthers, 2006; Koschmann, Stahl & Zemel, 2007). In an effort to investigate the organization of interactions across the dual-interaction spaces of the VMT environment, we consider the small group as the unit of analysis (Stahl, 2006), and we apply the methods of Ethnomethodology (EM) (Garfinkel, 1967; Livingston, 1986) and Conversation Analysis (CA) (Sacks, 1962/1995; ten Have, 1999) to conduct case studies of online group interaction. Our work is informed by studies of interaction mediated by online text-chat with similar methods (Garcia & Jacobs, 1998; O'Neill & Martin, 2003), although the availability of a shared drawing area and explicit support for deictic references in our online environment as well as our focus on mathematical practice significantly differentiate our study from theirs.

The goal of Conversation Analysis is to make explicit and describe the normally tacit commonsense understandings and procedures group members use to organize their conduct in particular interactional settings. Commonsense understandings and procedures are subjected to analytical scrutiny because they "enable actors to recognize and act on their real world circumstances, grasp the intentions and motivations of others, and achieve mutual understandings" (Goodwin & Heritage, 1990, p. 285). Group members' shared competencies in organizing their conduct not only allow them to produce their own actions, but also to interpret the actions of others (Garfinkel & Sacks, 1970). Since members enact these understandings and/or procedures in their situated actions, researchers can discover them through detailed analysis of members' sequentially organized conduct (Schegloff & Sacks, 1973).

We subjected our analysis of VMT data to intersubjective agreement by conducting numerous CA data sessions (ten Have, 1999). During the data sessions we used the VMT Replayer tool, which allows us to replay a VMT chat session as it unfolded in real time based on the timestamps of actions recorded in the log file. The order of actions—chat postings, whiteboard actions, awareness messages—we observe with the Replayer as researchers exactly matches the order of actions originally observed by the users. This property of the Replayer allows us to study the sequential unfolding of events during the entire chat session. In short, the VMT environment provides us a perspicuous setting in which the mathematical meaning-making process is *made visible* as a joint practical achievement of participants that is "observably and accountably embedded in collaborative activity" (Koschmann, 2001, p. 19).

## Analysis

The following sequence of drawing actions (Figures 2 to 6 below) is observed at the beginning of the very first session of a team in the VMT environment. Shortly after a greeting episode, one student, Davidcyl, begins to draw a set of squares on the shared whiteboard. He begins by drawing three squares that are aligned horizontally with respect to each other, which is made evident through his careful placement of the squares side by side (see Figure 2 below). Then he adds two more squares on top of the initial block of three, which introduces a second layer to the drawing. Finally, he adds a single square on top of the second level, which produces the stair-step shape displayed in the last frame of Figure 2. Note that he builds the pattern row-by-row here.

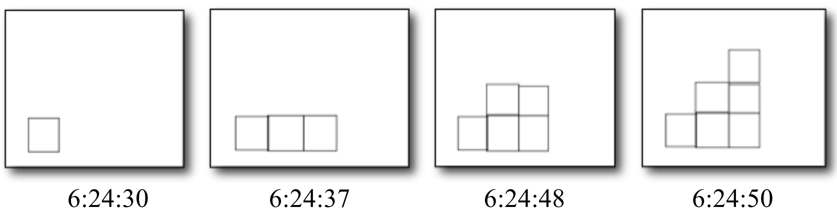
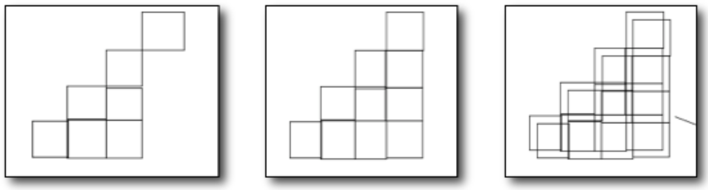



**Figure 2:** First stages of Davidcyl's drawing activity.

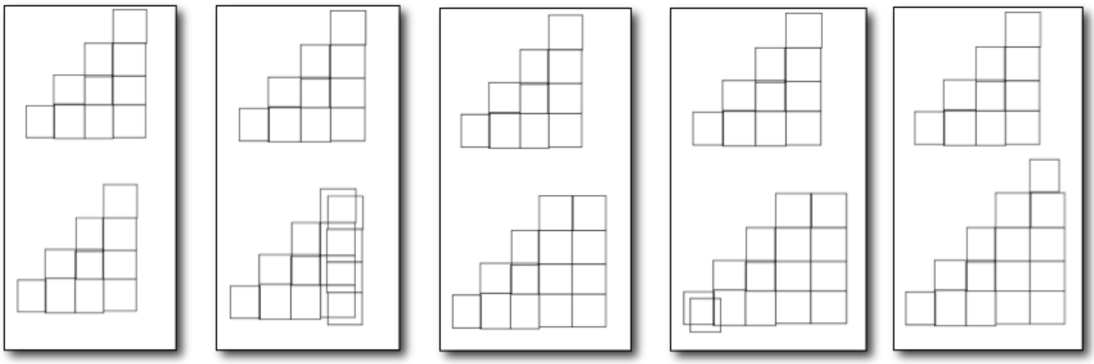
Next, Davidcyl starts adding a new column to the right of the drawing (see Figure 3). He introduces a new top level by adding a new square first, and then he adds 3 more squares that are aligned vertically with respect to each other and horizontally with respect to existing squares (see second frame in Figure 3). Then he produces a duplicate of this diagram by using the copy/paste feature of the whiteboard (see the last frame in Figure 3). Here, he builds the next iteration by adding a new column to the previous stage, starting the new column by making visible that it will be one square higher than the highest previous column.

Afterwards, Davidcyl moves the pasted drawing to an empty space below the copied diagram. As he did earlier, he adds a new column to the right of the prior stage to produce the next stage. This time he copies the entire 4$^{th}$ column, pastes a copy next to it, and then adds a single square on its top to complete the new stage (Figure 4). Next, Davidcyl produces another shape in a similar way by performing a copy/paste of his last drawing, moving the copy to the empty space below, and adding a new column to its right (see Figure 5). Yet, this time the squares of the new column are added one by one, which may be considered as an act of counting. In Figure 4, the new column is explicitly shown to be a copy of the highest column plus one square. In Figure 5, the number of squares in the new column are counted individually, possibly noting that there are N of them. The likelihood that the counting of the squares in the new column is related to the stage, N, of the pattern is grounded by Davidcyl's immediately subsequent reference to the diagrams as related to "*n=4,5,6*".



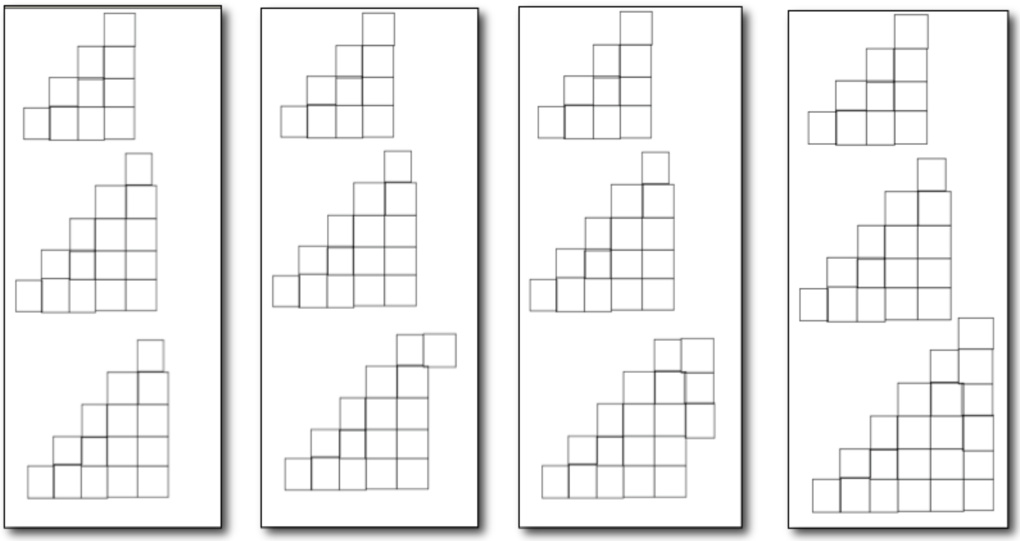
6:24:51 6:25:00 6:25:07

**Figure 3:** Davidcyl introduces the 4th column and pastes a copy of the whole shape.



6:25:13 6:25:42 6:25:45 6:25:47 6:25:52

**Figure 4:** Davidcyl uses copy/paste to produce the next stage of the pattern



6:26:04 6:26:10 6:26:14 6:26:20

**Figure 5:** Davidcyl's drawing of the 6th stage

Shortly after his last drawing action at 6:26:20, Davidcyl posts a chat message stating, "*ok I've drawn n=4,5,6*" at 6:26:25. Figure 6 shows the state of the interface at this moment. The "*ok*" at the beginning of the message could be read as some kind of a transition move (Beach, 1995). The next part "I've drawn" makes an explicit verbal reference to his recent (indicated by the use of past perfect tense) drawing actions. Finally, the expression "*n=4,5,6*" provides an *algebraic gloss* for the drawings, which specifies how those drawings should be seen or treated. Once read in relation to the task description, Davidcyl's recent actions across both spaces can be treated as a response to the first bullet under session 1, which states "Draw the pattern for N=4, N=5, and N=6 in the whiteboard" (see Figure 1 above for the task description). The discussion that immediately followed Davidcyl's drawings and his last chat statement is displayed in Table 1 below.

Davidcyl's posting at line 26 is stated as a declarative, so it can be read as a claim or assertion. The references to "*n*" (i.e., not to a particular stage like 2nd or 5th) invoke a variable as a gloss for referring to the features of the general pattern. Moreover, the use of the clause “more…than” suggests a comparison between two things, in particular the two cases indexed by the phrases “*nth pattern*” and *“(n-1)th pattern*” respectively. Hence, Davidcyl’s posting can be read as a claim about how the number of squares changes between the $(n-1)^{th}$ and $n^{th}$ stages of the pattern at hand. The two cases compared in the posting correspond to two consecutive stages of the staircase pattern. Davidcyl’s prior drawing work included similar transitions among pairs of particular stages. For instance, while he was drawing the $4^{th}$ stage, he added a column of 4 new squares to the right of the $3^{rd}$ stage. Hence, Davidcyl’s narrative uses the drawings for particular cases as a resource to index the properties of the general pattern, which is implicated in the regularity/organization projected by his prior drawing actions.

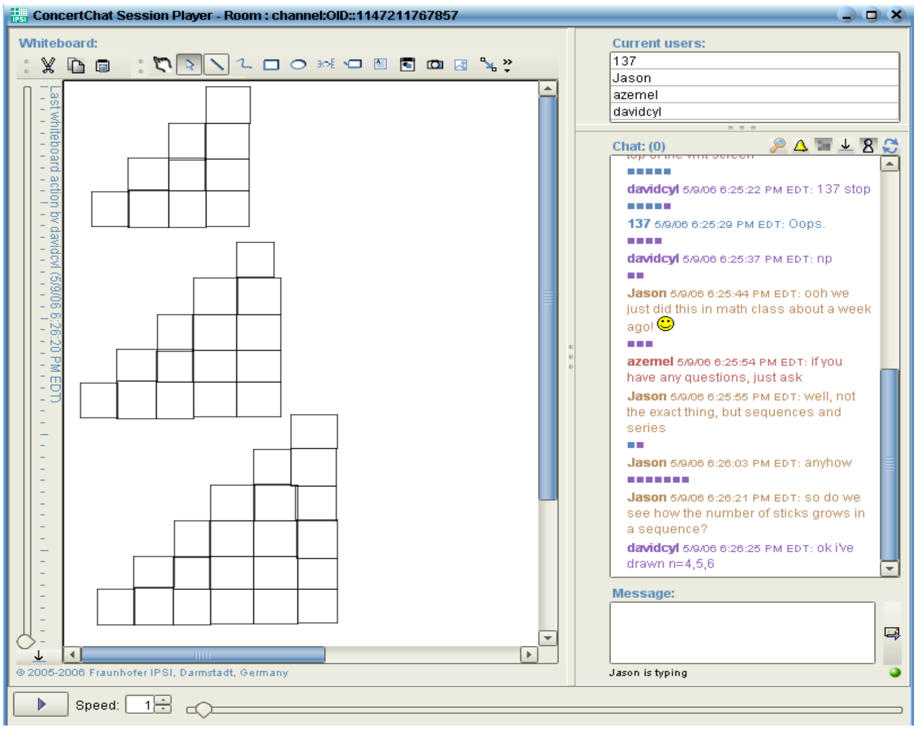



**Figure 6:** The state of the VMT environment when Davidcyl posted “ok I’ve drawn n=4,5,6” at 6:26:25.

**Table 1:** Chat discussion following the drawing activity

| Chat Index | Time Start Typing | Time of Posting | Author | Content |
|---|---|---|---|---|
| 26 | 18:27:13 | 18:27:32 | davidcyl | the nth pattern has n more squares than the (n-1)th pattern |
| | 18:27:30 | 18:27:47 | 137 | *[137 has fully erased the chat message]* |
| | 18:27:47 | 18:27:52 | 137 | *[137 has fully erased the chat message]* |
| 27 | 18:27:37 | 18:27:55 | davidcyl | basically it's 1+2+..+(n-1)+n for the number of squares in the nth pattern |
| | 18:27:57 | 18:27:57 | 137 | *[137 has fully erased the chat message]* |
| 28 | 18:28:02 | 18:28:16 | 137 | so n(n+1)/2 |
| 29 | 18:27:56 | 18:28:24 | davidcyl | and we can use the gaussian sum to determine the sum: n(1+n)/2 |
| 30 | 18:28:27 | 18:28:36 | davidcyl | 137 got it |

In the next line, Davidcyl elaborates on his description by providing a summation of integers that accounts for the number of squares required to form the $n^{th}$ stage. In particular, the expression “*1+2+..+(n-1)+n*” describes a method to count the squares that form the $n^{th}$ stage. Since Davidcyl made his orientation to columns explicit through his prior drawing work while he methodically added a new column to produce a next stage, this expression can be read as a *formulation* of his column-by-column counting work in algebraic form. In other words, Davidcyl achieves a (*narrative*) transition from the *visual* to the *algebraic*, which is informed by his methodic construction of specific stages of the staircase pattern that allowed him to isolate relevant components of the general pattern and derive a systematic counting method.

As Davidcyl composes a next posting, 137 posts a so-prefaced math expression at line 28, "*So n(n+1)/2*" that (a) shows 137 has been attending to the organization of Davidcyl’s ongoing exposition, (b) displays 137's recognition of the next problem-solving step projected by prior remarks, (c) offers an algebraic realization of the procedure described by Davidcyl, and (d) call on others to assess the relevance and validity of his claim. Davidcyl's message at line 29 (which is produced in parallel with line 28 as indicated by the typing times) is a more elaborate statement that identifies how his prior statements, if treated as a Gaussian sum, yield the same expression that 137 put forward at line 28 (viz. "n(n+1)/2"). Given that 137 anticipated Davidcyl's Gaussian sum, Davidcyl announces in the very next posting that "*137 got it,*" which recognizes the relevance of 137's posting at that particular moment in interaction, and treats 137's coordinated contribution as an act of understanding.

## Discussion

Given the characterization of deep mathematical understanding in the math education literature, methodic ways through which participants coordinate their actions across the whiteboard and chat spaces are of particular interest to our investigation of mathematical understanding or meaning making at the small-group level. The episode we analyzed above includes a situation where a user, who has been active in the whiteboard, moves on to the other interaction space and posts a message referring to his prior drawing work. The chat message sequentially followed the drawings, and hence presumed their availability as a shared referential resource, so that the interlocutors can make sense of what is possibly referred to by the indexical expression "*n=4,5,6*" included in the posting. Davidcyl's explicit orientation to timing or sequencing is further evidenced by his use of the past perfect tense and his temporal positioning of the message immediately after the final step of the drawing. Moreover, the chat posting reflexively gave further specificity to the prior drawing work by informing everyone that the diagrams should be seen as specific cases of the staircase pattern described in the task description. This suggests that temporal proximity among actions can serve as an interactional resource/cue for the participants to treat those actions in reference to each other, especially when the actions are performed across different interaction spaces. In short, Davidcyl has demonstrated a *method* that one can call *verbal referencing*, which is employed by VMT users when they need to communicate to each other that a narrative/symbolic account needs to be read in relation to a whiteboard object.

Davidcyl's use of the algebraic reference "*n=4,5,6*" at this moment in interaction is also informative in terms of respective limitations of each medium and their mutually constitutive function for communication. Davidcyl's chat message not only provided further specificity to the recently produced diagrams, but also marked or announced the completion of his drawing work. This is revealing in terms of the kinds of *illocutionary acts* (Austin, 1962) that can be achieved by users in this dual-media online environment. In particular, although a drawing and its production process may be available for all members to observe in the shared whiteboard, diagrams by themselves cannot fulfill the same kind of interactional functions achieved by text postings such as "asking a question" or "expressing agreement." In other words, whiteboard objects are made interactionally relevant through chat messages that either (a) project their production as a next action or (b) refer to already produced objects. This can also be seen as members' orientation to a limitation of this online environment as a communication platform; one can act only in one space at a given time, so it is not possible to perform a simultaneous narration of a drawing as one can do in a face-to-face setting. Therefore, each interaction space as a communicative medium seems to enable and/or hinder certain kinds of actions, i.e., they carry specific *communicative affordances* (Hutchby, 2001) for collaborative problem solving online.

The way Davidcyl has put some of the features of the whiteboard —like dragging and copy/paste—into use in the episode described above demonstrates some of its key affordances as a medium for producing shared drawings. In particular, we have observed how copying and pasting is used to avoid additional drawing effort, and how collections of objects are selected, dragged, and positioned to produce specific stages of a geometric pattern. Such possibilities for action are supported by the object-oriented design of the whiteboard. Davidcyl's drawing actions show that, as compared to other physical drawing media such as paper or blackboard, the electronic whiteboard affords unique possibilities for constructing and modifying shared mathematical diagrams in ways that have mathematical, collaborative, semantic and communicative power.

It is analytically significant that Davidcyl changed from building the pattern row-wise in Figure 2 to building it column-wise subsequently and that he "computed" the height of the new column in several different ways in Figures 3, 4 and 5. This indicates that he did not have an explicit solution "in his head"—a mental model that he just had to illustrate in the world with the whiteboard. Rather, he worked out the solution gradually through emergent whiteboard activities and his recognition of what appeared in the whiteboard. Significantly, the other group members could observe the same thing.

An important concern for our group-cognitive approach is to investigate how students make use of the technological features available to them to explore mathematical ideas in an online environment like VMT. Drawing features such as copy/paste, dragging, coloring, etc. are important affordances of the shared whiteboard not simply because of their respective advantages as compared to other drawing media. The mathematical significance of these features relies on the way single actions like copy/paste or dragging are sequentially organized as part of a broader drawing activity that aims towards constructing a shared mathematical artifact. For instance, through such a sequence of drawing actions Davidcyl demonstrated to us (as analysts) and to his peers (a) how to construct a stair-step pattern as a spatially organized assemblage of squares, and (b) how to derive a new stage of the stair-step pattern from a copy of the prior stage by adding a new column of squares to its right. Moreover, Davidcyl's engagement with the squares (rather than with the sticks that make up the squares) displays his explicit orientation to this particular aspect of the shared task (i.e., finding the number of squares at a given stage). Hence, the availability of these drawing actions as a sequence of changes unfolding in the shared visual space allows group members to witness the *reasoning* process embodied in the sequential and spatial organization of those actions. In other words, the sequentially unfolding details of the construction process provide specificity (and hence meaning) to the mathematical artifact that is being constructed.

Besides figuring out ways to connect their own actions across dual-interaction spaces, VMT users also coordinate their actions with the actions of their peers to be able to meaningfully participate in the ongoing discussion. The ways participants produce and deploy mathematical artifacts in the shared space implicate or inform what procedures and methods may be invoked next to produce other mathematical artifacts, or to modify existing ones as the discussion progresses towards a solution to the task at hand. For instance, 137's competent contribution to Davidcyl's sequentially unfolding line of reasoning in Table 1 shows that shared mathematical understanding at the group level is an interactional achievement that requires coordinated *co-construction* of mathematical artifacts. The *co* prefix for the term "co-construction" highlights the *intersubjective* nature of the mathematical artifacts produced during collaborative work; they are not mere mental constructs easily ascribed to certain individuals. As we have just observed in the excerpt above, *intersubjectivity* is evidenced in the ways participants organize their actions to display their relevance to prior actions. 137's anticipation and production of the next relevant step in the joint problem-solving effort serves as strong evidence of mutual understanding between him and Davidcyl. Moreover, the term *construction* signals that mathematical artifacts are not simply passed down by the mathematical culture as ready-made Platonic entities external to the group. Once enacted in group discourse, culturally transmitted artifacts such as "Gaussian Sum" need to be made sense of and appropriated in relation to the task at hand. Hence, our use of the combined term *co-construction* implies an interactional process of sense making by a group of students—even in an excerpt like the present one in which one individual takes an extended turn in the group discourse to develop a complex presentation. The fact that it is a visible construction worked out in collaborative media and designed for reception by others makes it a co-construction from which the speaker is as likely to learn as are the other group members.

When co-construction takes place in an online environment like a chat tool, the construction process must take place through observable interactions within technical media. This requires student groups to invent, adapt or appropriate methods to co-construct mathematical artifacts. It also makes it possible for them to explicitly reflect on the *persistent traces* of their co-constructions by investigating the persistent content provided by the technology. Therefore, the persistent nature of actions provides the necessary infrastructure for joint action, and hence is a key affordance of CSCL environments like VMT, where actors work at a distance in a disembodied environment. In addition, the persistent records of interactions also allow researchers to analyze the co-construction process as it unfolded in real-time, as this paper demonstrates.

Through similar case studies of other VMT sessions, we observed that students make use of additional resources (such as the explicit referencing tool, locational pronouns, color names in chat) to methodically achieve referential relationships between shared diagrams and chat messages (Cakir, Zemel & Stahl, 2009). Chat postings use a broad and sophisticated array of such methods to refer to matters constructed graphically. Due to their recurrent appearance as a practical concern for the participants in this dual-media online environment, we refer to the collection of these methods as *referential practices*. Referential practices are of particular importance to the study of mathematical understanding as a group-cognitive phenomenon, because they are enacted in circumstances where participants explicitly orient to the task of achieving relationships between the textual and graphical contributions that they have been exchanging online—a phenomenon that is given significance in the math education literature as characterizing deep mathematical understanding. Likewise, one can use the term *representational practices* to refer to the spatial and temporal organization of whiteboard actions that produce shared diagrams, which simultaneously give further specificity to the mathematical artifacts that the team has been working on—e.g., Davidcyl's methodical sequencing of copy/paste operations to indicate growth patterns. Through referential and representational practices, participants co-construct mathematical artifacts that reify mathematical understandings. The understanding or meaning is not simply located inside students' individual brains or in the chat/drawing artifacts themselves. The meaning is embodied in the sequentially organized and coordinated actions through which those artifacts were co-constructed. To sum up, group referential and representational practices play a key role in the ways mathematical artifacts are (a) appropriated by active teams from historically developed cultural tools, and (b) emergent from ways of communicating and symbolizing within local collectivities as shared, meaningful resources for mathematical discourse, collaborative learning and group understanding.

## References


Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

Beach, W. A. (1995). Conversation analysis: "Okay" as a clue for understanding consequentiality. In S. J. Sigman (Ed.), *The consequentiality of communication* (pp. 121- 162). Hillsdale, NJ: LEA.

Cakir, M. P., Zemel, A., & Stahl, G. (2009). *The joint organization of interaction within a multimodal CSCL medium.* International Journal of Computer-Supported Collaborative Learning, 4(2), 155-190.

Dillenbourg, P., & Traum, D. (2006). Sharing Solutions: Persistence and Grounding in Multimodal Collaborative Problem Solving. *The Journal of the Learning Sciences, 15*(1), 121-151.

Dohn, N. B. (2009). Affordances revisited: Articulating a Merleau-Pontian view. *International Journal of Computer-Supported Collaborative Learning, 4*(2), 151-170.

Garcia, A., & Jacobs, J. B. (1998). The interactional organization of computer mediated communication in the college classroom. *Qualitative Sociology, 21*(3), 299-317.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

Garfinkel, H., & Sacks, H. (1970). On formal structures of practical actions. In J. Mckinney & E. Tirvakian (Eds.), *Theoretical sociology: Perspectives and developments* (pp. 337-366). New York: Appleton.

Goodwin, C., & Heritage, J. (1990). Conversation Analysis. *Annual Review of Anthropology, 19*, 283-307.

Healy, L., & Hoyles, C. (1999). Visual and symbolic reasoning in mathematics: Making connections with computers. *Mathematical Thinking and Learning, 1*(1), 59-84.

Hiebert, J., & Wearne, D. (1996). Instruction, Understanding, and Skill in Multidigit Addition and Subtraction. *Cognition and Instruction, 14*(3), 251-283.

Hutchby, I. (2001). *Conversation and technology: From the telephone to the internet*. Cambridge, UK: Polity.

Kaput, J. (1998). Representations, Inscriptions, Descriptions and Learning: A Kaleidescope of Windows. *Journal of Mathematical Behavior, 17*(2), 265-281.

Koschmann, T. (2001). Revisiting the paradigms of instructional technology. In G. Kennedy, M. Keppell, C. McNaught & T. Petrovic (Eds.), *Meeting at the crossroads. Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education* (pp. 15-22).

Koschmann, T., Stahl, G., & Zemel, A. (2007). The video analyst's manifesto. In R. Goldman, R. Pea, B. Barron & S. Derry (Eds.), *Video research in the learning sciences*. Mahwah, NJ: LEA.

Livingston, E. (1987). *Making sense of ethnomethodology*. New York, NY: Routledge.

Moss, J., & Beatty, R. (2006). Knowledge building in mathematics: Supporting collaborative learning in pattern problems. *International Journal of Computer-Supported Collaborative Learning, 1*(4), 441-465.

Mühlpfordt, M., & Wessner, M. (2005). Explicit Referencing in Chat Supports Collaborative Learning. In T. Koschmann, D. Suthers & T. Chan (Eds.), *Proceedings of the 2005 conference on Computer support for collaborative learning (CSCL): the next 10 years!* (pp. 460-469). Taipei, Taiwan: ISLS.

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

O'Neil, J., & Martin, D. (2003). Text chat in action. In K. Schmidt & M. Pendergast (Eds.), *GROUP '03: Proceedings of the ACM SIGGROUP on Supporting group work* (pp. 40-49). New York, NY: ACM.

Sacks, H. (1962/1995). *Lectures on Conversation*. Oxford, UK: Blackwell.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 7*, 289-327.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. New York, NY: Cambridge University Press.

Skemp, R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*, 20-26.

Soller, A. (2004). Understanding knowledge sharing breakdowns: A meeting of quantitative and qualitative minds. *Journal of Computer Assisted Learning, 20*, 212-223.

Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge*. Cambridge, MA: MIT Press.

Stahl, G. (Ed.). (2009). *Studying virtual math teams*. New York, NY: Springer.

Stahl, G., Koschmann, T., & Suthers, D. D. (2006). Computer-Supported Collaborative Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 409-426). New York: CUP.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning (ijCSCL), 1*(3), 315-337.

Suthers, D., Connelly, J., Lesgold, A., Paolucci, M., et al. (2001). Representational and Advisory Guidance for Students Learning Scientific Inquiry. In K. D. Forbus & P. J. Feltovich (Eds.), *Smart Machines in Education: The Coming Revolution in Educational Technology* (pp. 7-35). Menlo Park: AAAI Press.

ten Have, P. (1999). *Doing conversation analysis: A practical guide*. Thousand Oaks, CA: Sage Publications.

Watson, A., & Mason, J. (2005). *Mathematics as a constructive activity: Learners generating examples*. Mahwah, NJ: Lawrence Erlbaum Associates.

# Equity in Scaling Up SimCalc: Investigating Differences in Student Learning and Classroom Implementation

Jeremy Roschelle, SRI International, jeremy.roschelle@sri.com
Jessica Pierson, San Diego State University, jpierson@mail.sdsu.edu
Susan Empson, University of Texas at Austin, empson@mail.utexas.edu
Nicole Shechtman, SRI International, nicole.shechtman@sri.com
Margie Dunn, Rutgers University, Newark, dunn@cs.rutgers.edu
Deborah Tatar, Virgrnia Tech, tatar@vt.edu

**Abstract:** The Scaling Up SimCalc project implemented three large-scale studies designed to evaluate the impact of a Learning Sciences-based replacement unit targeting student learning of advanced middle school mathematics. Strong main effects in each study consistently showed the approach to be effective in enabling a wide variety of teachers in a diversity of settings to extend student learning to more advanced mathematics. In this paper, we take a closer look at *equity* of learning, the extent to which students within and across subgroups and classrooms had comparable learning gains. We describe four different patterns of equity and inequity that have emerged in our hierarchical linear modelling and in-depth case studies. We then present a number of conjectures about possible mechanisms by which opportunity to learn may be unevenly distributed between and within classrooms engaged in the same replacement unit. We discuss implications for instructional design, research methodology, and Learning Sciences theory.

## Introduction

In an on-going research program, we have been studying large-scale implementations of the Learning Sciences-based SimCalc approach to integrating representational technology, curriculum workbooks, and teacher professional development (see Roschelle et al., in press). The SimCalc program's slogan, "Democratizing Access to the Mathematics of Change and Variation," suggests a significant concern with equity (Kaput, 1994). Indeed, the design experiment research preceding our current large-scale work focused deliberately on giving diverse students in traditionally underperforming settings an opportunity to learn more advanced mathematics concepts. Hence, design experiments were conducted in diverse and economically challenged places such as Fall River, Massachusetts; Syracuse, New York; Newark, New Jersey; and San Diego, California (Roschelle, Tatar, & Kaput, 2008). This concern with equity was also reflected in the subsequent large-scale experiments, which occurred in Texas. We deliberately recruited schools in multiple regions of Texas to reflect that state's considerable economic, linguistic, and ethnic diversity (Roschelle et al., in press).

Equity has been an elusive goal in the reform of mathematics education. For example, prior research found that middle- to higher-achieving students benefitted from instruction that focused on integrating concepts and procedures, but lower-achieving students did not (Baxter, Woodward, & Olson, 2001); and achievement gaps in mathematics between white students and non-white students persist (Neal, 2005). Furthermore, there are competing definitions of equity (Lynch, 2000). Does it involve providing high-quality resources for all classrooms, ensuring that learning activities have ways for all students to participate meaningfully, closing achievement gaps, or perhaps some combination of these things? In this paper, we explore equity in the context of a large-scale implementation of a Learning Sciences–style approach, one integrating dynamic representations and visualizations to address learning of more advanced mathematical constructs.

After providing a brief overview of the program of research, we examine equity in the Scaling Up SimCalc Project. First, we describe four patterns of equity and inequity we have observed related to student achievement. We use a variety of indicators across multiple units of analysis. Second, we consider a variety of conjectured explanations for inequitable classrooms that are emerging in ongoing data analyses. Two highlighted conjectures are blocked access to learning resources and teacher talk that is not responsive to students and does not place a high level of intellectual demand on students. Other possible explanations for inequity are also considered.

## Overview of the Scaling Up SimCalc Project

The Scaling Up SimCalc Project (Roschelle et al., in press) implemented two randomized controlled experiments (with one embedded quasi-experiment) designed to address the broad research question, "Can a wide variety of teachers use an integration of technology, curriculum, and professional development to increase student learning of complex and conceptually difficult mathematics?" There were two interventions, one for seventh grade and one for eighth grade. Each intervention integrated the representational technology SimCalc

MathWorlds, curriculum workbooks, and teacher professional development organized around a 2-3 week replacement unit on rate, proportionality, and linear function. The replacement units incorporated the following hallmarks of the SimCalc approach to the mathematics of change and variation:

1. Anchoring students' efforts to make sense of conceptually rich mathematics in their experience of familiar motions, which are portrayed as computer animations.
2. Engaging students in activities to make and analyze graphs that control animations.
3. Introducing piecewise linear functions as models of everyday situations with changing rates.
4. Connecting students' mathematical understanding of rate and proportionality across key mathematical representations (algebraic expressions, tables, graphs) and familiar representations (narrative stories and animations of motion).
5. Structuring pedagogy around a cycle that asks students to make and compare their predictions.

The SimCalc MathWorlds software provides a "representational infrastructure" (Kaput, Hegedus, & Lesh, 2007) that is central to enabling this approach. Most distinctively, the software presents animations of motion. Students can control the motions of animated characters by building and editing mathematical functions in either graphical or algebraic forms. The program developers view student use of the software and teacher explanations and teacher-led discussions as complementary activities.

In the scaling research, across over 100 seventh- and eighth-grade classrooms throughout the state of Texas, the main effects of the treatment in the three main studies were positive, with student-level effect sizes of .63, .50 and .56; classrooms that used a SimCalc replacement unit had students who learned more advanced mathematics.

## Four Patterns of Equity and Inequity

Programs can be more or less equitable at a variety of levels. While overall there were robust learning gains, here we drill down more deeply at demographic and regional levels, as well as the classroom level, to examine three patterns of equity and inequity.

### 1. Equity at the Demographic and Regional Levels

As Figure 1 shows, at the highest level of aggregation, we found learning gains to be equitable across major demographic and regional groups. In the treatment group, we found no statistically significant differences in gains between gender or ethnic groups, or among students in schools serving different levels of socioeconomic populations. We also found no differences among geographic regions. This is particularly notable for "Region 1" of Texas, which covers the Rio Grande Valley along the Mexican border and has a high-poverty population mainly of Mexican descent. The effects of SimCalc were similar in this region to those in other regions sampled in the study (e.g., the more cosmopolitan regions near the state capital, Austin). These findings of equity were replicated in the Eighth-Grade study (see Roschelle et al., in press).

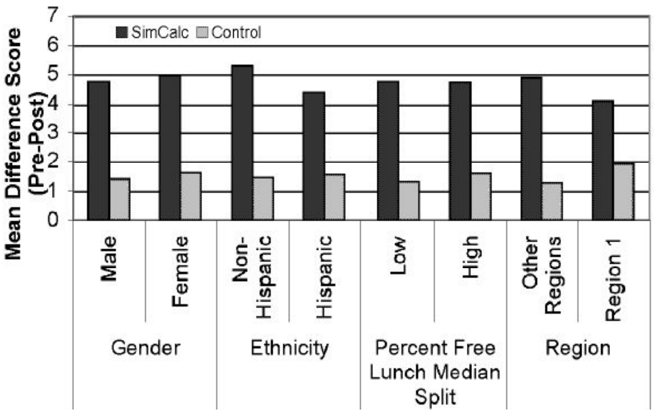



Figure 1. Mean student-learning gains by subpopulation group (Seventh-Grade Study).

### 2. Within-Classroom Spread of Learning Gains

While equity of learning occurred at this highly aggregated level, at the classroom level, do students learn equitably? To begin to examine this issue, we discuss the work of Empson, Greenstein, Maldonado, and Roschelle (2009) who report comparative case studies of three teachers who participated in the Seventh-Grade Study and taught with the SimCalc replacement unit. Two of the teachers, "Ms. Garfield" and "Ms. Driver" (pseudonyms), had mean classroom gains at or above the average mean gain found in the treatment group. The third teacher, "Mr. Simmons," had a mean classroom gain that was lower than the average by a third of a standard deviation. Note that mathematical knowledge for teaching (MKT) was not a predictor of student learning in these three cases. Mr. Simmons scored the highest of the three on our MKT assessment. Indeed, he was a former computer specialist. In a more detailed analysis of MKT we argue that MKT is just one of a variety of resources in the classroom instructional system (Shechtman, Roschelle, Haertel, & Knudsen, in press).

To investigate equity, we examine two aspects of learning in the classroom: overall gains and variation between students in gains. As Figure 2 shows, while the median gain in Mr. Simmons' class was relatively low, there was also a notable spread in achievement among his students. A few students performed very well in this classroom, but many students did much more poorly. In the two other cases, gains were both higher and less variable (as evidenced by less spread) among students. If one defines equity as equal learning gains for *all* students, then Mr. Simmons' classroom illustrates inequity.

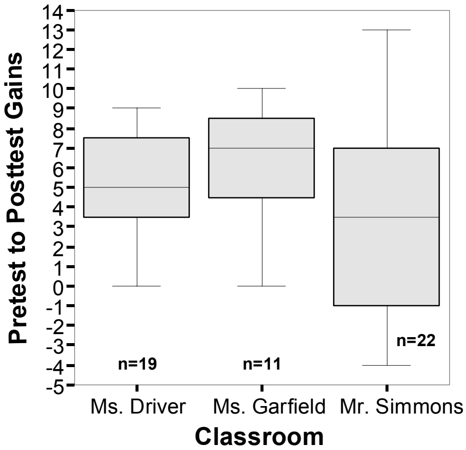



Figure 2. Boxplots of distributions of student gains on the 30-item assessment across the case study classrooms.

## 3. Within-Classroom Pretest-Posttest Slopes Tend to Be Steeper in Lower-Achieving Classrooms

Another approach to investigating equity is by examining the relationship within a classroom between students' prior knowledge (i.e., their pretest scores) and their learning gains. Pierson's (2008) dissertation investigated some of these patterns across classrooms. To illustrate, Figure 3 shows student achievement data in which each student is represented by an ordered pair where the x-coordinate represents his or her pretest score, and the y-coordinate represents the corresponding posttest score. The figure displays such data and best-fit regression lines for two classrooms. The variable of interest with respect to equity is the pretest-posttest achievement slope. Some classrooms have steeper slopes and some have flatter slopes. In other words, pretest scores are stronger determinants of achievement in some classrooms. Some claim that more equitable classrooms are those in which the pretest-posttest slope is closer to 0. In other words, *all* students achieve equally regardless of their prior knowledge.

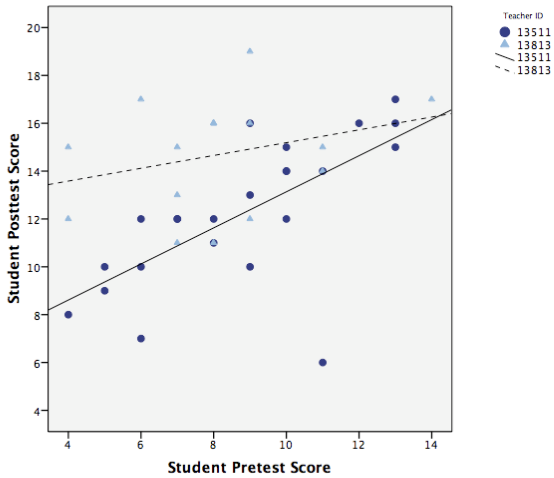



Figure 3. Scatter plot and regression lines for student pretest and student posttest scores in two SimCalc classrooms.

Pierson found in her HLM model a negative relationship between the mean achievement and pretest-posttest slopes at the class level ( $\tau_{01} = -.4309$). In other words, the higher the class posttest average, the smaller the pre-post achievement slope.[1] Classrooms that learned more overall (i.e., had higher average posttest scores) also learned more equitably.

### 4. The "Low-Low" Effect

We also investigated equity of learning across students' prior achievement levels (both as rated by their teachers and indicated by their pretest scores). We asked teachers to rate their students as low, medium, or high achieving prior to using the SimCalc replacement unit. Analyses found that the teachers' rating accurately predicted outcomes: students designated by their teacher as low-achieving learned less. Of course, the opposite effect would be desirable: if teachers knew which students were low-performing, it would be preferable for those students to receive *more* resources and guidance so that they could learn more.

More broadly, we are investigating a "low-low effect," designated as such because it appeared for low-ranked students in classrooms with low pretest means. We found this effect in both our Seventh- and Eighth-Grade studies, but for brevity we only present the Seventh-Grade data in Figure 4. Three graphs are presented side by side, representing classrooms with low-, medium-, and high-mean pretest scores (based on a tertile split across the treatment group). Each graph plots the teachers' categorization of their students (as low, medium, or high achievers) against the mean of those students' gain scores. The students in the low classroom have a particular pattern: low students in low classrooms had the lowest gains.

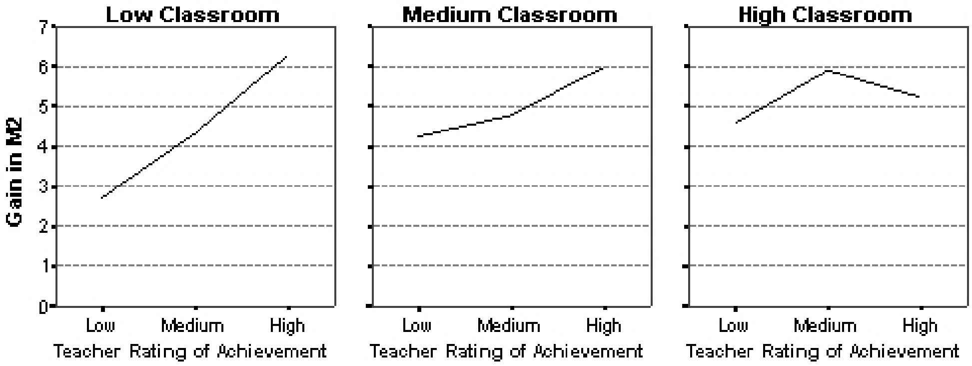



Figure 4. The "low-low effect" in the Seventh-Grade treatment group.

We have built an HLM of this phenomenon, and it is statistically significant in both our Seventh- and Eighth-grade studies. For each study, we ran an HLM of student gains, nesting students within classrooms. We then constructed the model using classroom mean pretest scores as a predictor at the classroom level and indicator variables of teacher-rated student achievement level (low and middle levels, with high level as the reference category) as predictors at the student level. To test for the low-low effect, we examined the coefficient for the interaction between classroom pretest and the indicator for low-achieving student. This coefficient was positive and significant in both the seventh-grade study ($\beta = 0.29$, $p < .05$) and the eighth-grade study ($\beta = 0.40$, $p < .001$), indicating the presence of the low-low effect. We also tested the same model on control classrooms where SimCalc materials were not used and found no presence of low-low effect. This suggests that the low-low effect is specific to the SimCalc materials.

We are in the process of investigating the low-low effect more deeply to examine possible alternative explanations for this phenomenon, including the possibility that it is an artifact of some kind. For example, it could be that only low classrooms have students with extremely low pretest scores. And it could be that students need a certain level of knowledge at baseline in order to engage meaningfully with the curriculum, or that teachers need certain additional skills to use these materials that engage lower-achieving students. For present purposes, the important point is that we see this as an indication of more and less equitable implementations of the SimCalc replacement unit.

## Conjectured Mechanisms of Within-Classroom Equity and Inequity

While our research was not designed to pinpoint particular mechanisms that produce more or less equitable classroom implementations, empirical analyses have produced a number of candidate explanations.

### 1. Blocked Access to Learning Resources during Instruction

One conjecture that has emerged in Empson et al.'s (2009) case studies is that some teachers may block learning resources from students, thus causing inequitable learning opportunities. This grows out of the idea that

SimCalc can work in multiple configurations of learning resources. Full class discussion around a shared simulation display, guided by a knowledgeable teacher, can be one kind of resource. Individual student activities with the software, guided by a workbook, can be another kind of resource. Empson and her team found that Ms. Garfield and Ms. Driver each had classroom implementations which featured different configurations of resources. What was distinctive about Mr. Simmons was that he blocked access to both full class discussion and individual student activities for many students. For example, Mr. Simmons concentrated whole-group discussion on interaction with a small subset of vocal boys, who made over 10 times more contributions than rest of the class. He also curtailed students' interactions with SimCalc by specifying each step to take in running the simulation and by requiring students to frequently turn away from the computers. Blocking access to resources in these ways decreased students' opportunities to engage meaningfully with the content and to make substantive connections on their own.

The pattern of blocking access was also prominent in Dunn's case studies in the Eighth-Grade study (Dunn, 2009). One teacher, "Marilyn," had particularly low student gain scores. She was on the low end of the range in total time allocated to teaching with the SimCalc materials, and she skipped more material than the other teachers. In particular, Marilyn seemed to have a pattern of beginning each day's class with a new workbook lesson, regardless of whether there had been sufficient time to complete the previous workbook lesson in the previous class sessions. As the workbook was designed to build ideas sequentially, skipping material in this way could block coherent development of mathematical ideas. Marilyn also was at the high end of the spectrum for the time spent introducing students to ideas via lecture, with a tendency to "walk" students through significant portions of each workbook lesson, in most cases at the beginning of class. This preempted independent student interaction with these portions, and compromised the time available for subsequent autonomous student work and class discussion/teacher feedback on this later work. This uneven allocation of time and guidance blocked consistent access and, once again due to the sequential design of the materials, may also have blocked coherent development of the unit's mathematical ideas.

The complement to the idea of blocked access is the idea of multiple learning resources and multiple configurations of learning opportunities. It could be that a key design principle for equity on a large scale is to provide multiple means for students to engage with the big ideas within the same intervention.

## 2. Favorable Teacher Discourse Moves

A second major conjecture examines the role of particular types of teacher talk in the classroom – responsiveness to student ideas and how much intellectual work they demand of students. Pierson's (2008) dissertation used HLM to examine teacher discourse across 13 teachers who taught using SimCalc materials. In particular, Pierson analyzed videotapes and transcripts of the same SimCalc lesson for all teachers. She developed the ideas that teachers (a) can be more or less responsive[2] to their students and (b) can demand more or less intellectual work[3] from their students. Using a slightly different HLM model than that described above for the low-low effect, she ran her analysis on student posttest achievement scores, nesting students within classrooms. She used classroom discourse variables as predictors at the classroom level and pretest achievement scores as a predictor at the student level.

There were two findings relevant to this discussion of equity. First, Pierson found that both high responsiveness and high intellectual work correlate with classroom-level *gains*. Second, these factors also correlated with the pretest-posttest slope (see discussion above). Specifically, Pierson found that an *increase* of 10% in highly responsive moves corresponded to a *decrease* in the class's pretest-posttest slope of .29. Interestingly, responsiveness and intellectual work seem to be factors in more equitable classrooms, as shown by a decrease in the impact of pretest scores on posttest achievement.

Although Pierson's study showed good evidence *that* high levels of responsiveness and intellectual work have a positive relationship with more equitable patterns of achievement, we should consider *how* these discursive constructs might be related to equity. We hypothesize the following three more refined mechanisms as possible links between equitable achievement and patterns of discourse.

First, discussions high in responsiveness and intellectual work can act as formative assessments, which research indicates is positively related to student learning (Fennema et al., 1996; Wiliam, Lee, Harrison, & Black, 2004). These types of discourse moves afford teachers opportunities to gather evidence of student understanding in order to adjust and customize instruction, resulting in better learning opportunities for everyone.

Second, repeated patterns of responsiveness and intellectual work in classroom discourse socialize students into specific ways of being. These normative discourse patterns communicate expectations for how to engage with one another (listening to, learning from, and critiquing others' ideas); expectations about learning (mistakes are a natural process in learning; and learning is about thinking, not remembering); and beliefs about one's own abilities. Students, particularly low-achieving students, are likely to develop increased confidence, motivation, persistence, and more positive mathematical identities as their ideas are valued, taken up, and extended.

Third, high levels of intellectual work and responsiveness create classroom cultures that require students to routinely work on cognitively demanding tasks that require argumentation, justification, and cognitive struggle. The teachers' role here is critical in that they must uphold the expectation that all students will generate their own solutions and explain them, position students as capable problem solvers, and carefully orchestrate student involvement in classroom discussions. Creating this type of learning community might benefit students with lower levels of prior knowledge reduce that gap and achieve at levels comparable with their peers at the top of the class.

## 3. Emerging Conjectures in Other SimCalc Case Studies

Other case studies conducted in the context of the Scaling Up SimCalc studies are still underway:

- Work by Michelle McLeese and Deborah Tatar considers the social climate of the classroom. This study is exploring the idea that certain forms of laughter signal a more positive social climate and contribute to better learning outcomes.
- Another study, by Meg Kurdziolek and Deborah Tatar considers the configuration of technology in the classroom. For example, they observed SimCalc implementations in a computer lab, with a laptop cart in the teacher's regular classroom, and with only one computer per classroom. Surprisingly, the implementation in the classroom with only one computer was successful; the teacher found ways to actively involve all students despite the resource limitation. They also have found instances where a classroom has more computers and students work in pairs, but a more dominant partner blocks engagement of a less dominant partner. Likewise, both Dunn (2009) and Empson et al. (2009) observed classrooms with a sufficient number of computers where student activity with the computer was routinely interrupted or blocked by the teacher, leading to poor outcomes. At the very least, this suggests that analysis of equity in relationship to the numbers of computers in a classroom needs to take into account pedagogical adaptations and patterns of student interaction when they work in small groups.
- We are also analyzing workbook materials collected in diverse classrooms, containing student work. Preliminary analysis of these workbooks has found strong correlations from students' pretest scores to the completeness and correctness of their workbooks. This may suggest an equity concern that students who begin with higher math scores are better prepared to use the workbooks. Furthermore, because completing relevant portions of the student workbook is correlated with student gains, providing more support to students who are falling behind in completing the workbook might contribute to more equitable outcomes.

As these additional studies mature, they may contribute more nuance to our concepts of access to learning resources and favorable teacher discourse moves, or may contribute additional equity-relevant constructs.

## Discussion

Equity is an important policy objective and often an important goal in the design of interventions resulting from Learning Science theories and design experiments.

By presenting results from the Scaling Up SimCalc work, we have highlighted a few key ideas:

1. Equity and inequity occur simultaneously on different levels and scales; equity may be present at some levels and absent in others.
2. Classroom implementations of the same intervention can be more or less equitable.
3. A range of mechanisms may contribute to equity and inequities in particular classroom interventions.

With regard to the last point, we identified two relevant mechanisms. First, even the best designed learning resources are not useful if students' access to those resources is blocked. Second, when teachers are responsive to student ideas and engage students in doing public intellectual work, all students seem to benefit, not just those students that the teacher is interacting with. However, findings pertaining to students in the low-low effects cases suggest that additional teacher development may be needed in order for instruction to support lower-achieving students' productive engagement with the content. We need to know more about how the teaching knowledge and practices are involved in being responsive and posing questions with high intellectual demand for these students.

These observations may benefit *design*: if we become aware of different sources of inequity, we may be able to design features into the intervention to mitigate some. For example, we could give teachers more guidance about making SimCalc resources available to students in multiple configurations and giving them time to use the materials autonomously.

The observations may also benefit *method.* We suggest that a mix of large-scale experimental designs with embedded case studies is a particularly fruitful method to ferret out issues of inequity, for a few reasons. Large-scale studies can be designed with a structure that captures data at the multiple levels at which equity and inequity can occur. For example, our designs considered the regional, school, classroom, and student levels. Hierarchical linear modeling can highlight cross-level patterns of inequity; some classrooms may have flatter slopes correlating student pretests to gains (i.e., more equitable outcomes), while other classrooms have steeper slopes (e.g., the rich get richer). Case studies, however, are also important for detecting mechanisms of inequity. For instance, it was far from obvious to us at the beginning of the experiments that we might see one-computer-per-classroom implementations that were more equitable than implementations with multiple computers.

We also suggest that this research could lead to fruitful paths in Learning Science *theory*. Although much Learning Science research looks at discourse, less looks at the implications of various classroom discourse strategies for equity. The finding that responsiveness and intellectual demand may help level the playing field in the classroom suggests that the subject of teacher discourse moves is worthy of further investigation.

Overall, we are encouraged about the range of settings in which we found equitable outcomes and energized by the potential to better pinpoint sources of inequity and to better address them in future design iterations.

## Endnotes

[1] A ceiling effect is unlikely because highest average pretest score for any class in Pierson's study was 9.00 points/problems (out of 18 possible) with the average pretest score across all classes being 5.877 points (*s* = 2.28).

[2] *Responsiveness* is an attempt to understand what another is thinking displayed in how one builds, questions, clarifies, takes up, or probes that which another says.

[3] *Intellectual work* reflects the cognitive work requested from students with a given turn of talk. High levels of intellectual work extend thinking and include discursive moves, such as providing justifications, examples, conjectures, explanations, and challenges; making connections across representations; generating problems and scenarios (contextualizing); or requesting these activities from students.

## References


Baxter, J., Woodward, J. & Olson, D. (2001). Effects of reform-based mathematics instruction on low achievers in five third-grade classrooms. *The Elementary School Journal, 101,* 529-547.

Dunn, M. (2009). *Investigating variation in teaching with technology-rich intervention: What matters in teaching and teacher training at scale?* Unpublished Ed.D., Rutgers University, New Brunswick, NJ.

Empson, S. B., Greenstein, S., Maldonado, L., & Roschelle J. (2009). Scaling up innovative mathematics in the middle grades: Case studies of "good enough" enactments. Manuscript submitted for publication.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi. L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403-434.

Kaput, J. (1994). Democratizing access to calculus: New routes using old roots. In A. Schoenfeld (Ed.), *Mathematical thinking and problem solving* (pp. 77-155). Hillsdale, NJ: Erlbaum.

Kaput, J., Hegedus, S., & Lesh, R. (2007). Technology becoming infrastructural in mathematics education. In R. Lesh, E. Hamilton & J. Kaput (Eds.), *Foundations for the Future in Mathematics Education* (pp. 173-192). Mahwah, NJ: Lawrence Erlbaum Associates.

Lynch, S. J. (2000). *Equity and science education reform.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Neal, D. (2005). Why has Black-White skill convergence stopped? Retrieved June 2005 from http://home.uchicago.edu/~n9na/.

Pierson, J. (2008). *The relationship between patterns of classroom discourse and mathematics learning.* Unpublished doctoral dissertation, The University of Texas at Austin.

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J. & Gallagher, L. (in press). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal.*

Roschelle, J., Tatar, D., & Kaput, J. (2008). Getting to scale with innovations that deeply restructure how students come to know mathematics. In A. E. Kelly, R. Lesh & J. Y. Baek (Eds.), *Handbook of design research methods in education* (pp. 369-395). New York: Routledge.

Roschelle, J., Tatar, D., Shechtman, N., & Knudsen, J. (2008). The role of scaling up research in designing for and evaluating robustness. *Educational Studies in Mathematics, 68*, 149-170.

Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (in press). Investigating links from teacher knowledge, to classroom practice, to student learning in the instruction system of the middle school mathematics classroom. *Cognition and Instruction.*

Wiliam, D., Lee, C., Harrison, C. & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49-65.

## Acknowledgments
This material is based in part upon work supported by the National Science Foundation under Grant Number 04-37861. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

# Distributed Creativity Within a Community of Student Instructional Designers

**Abstract:** This study explored the development of a Communities of Innovation (COI) framework for understanding distributed creativity within a community of graduate student instructional designers. After presenting distributed creativity as a theory of collaborative creativity based on established principles of distributed cognition, I present the Communities of Innovation framework as a potential representation of distributed creativity. I then discuss a study where phenomenological interviewing (Seidman, 2006) and Critical Incident Technique (Flanagan, 1952), were used to explore the experiences of four members of a graduate community of designers with many characteristics emblematic of COIs. Findings included evidence for the inclusion of some aspects of the proposed COI framework. In addition, I identify challenges and recommendations to establishing a COI within a graduate educational setting and possible new directions for research using a variety of different methods to better understand the nature of COIs and how to effectively develop them.

## Understanding Distributed Creativity

Creativity research has often replicated and adapted the theoretical frameworks and research methodologies of cognitive science. For example, creativity researchers have drawn on cognitive principles such as knowledge and memory structures, representations, interference, and so on (for example, see Smith, Ward, & Finke, 1995) in an effort to try and understand the nature of human creativity. However, when cognitive scientists began exploring the potential distribution of cognition across multiple people, creativity researchers largely remained behind. While some writers explored the relationship between individual creativity and the overall system (Csikszentmihalyi, 1999) or the group or workplace climate (Amabile, Conti, Coon, Lazenby, & Herron, 1996; Anderson & West, 1996), few have studied the creative process in distributive terms. More recently, there has been a stronger emphasis on group, or collaborative, creativity—a trend offering great promise for traits increasingly valued in our society (West, 2009).

In this paper presentation, I first discuss key principles necessary for understanding the theory of distributed cognition. I then extend Sawyer and DeZutter's (2009) application of these principles to the concept of distributed creativity and discuss the Communities of Innovation framework as an example of distributed creativity. Finally, I discuss findings from a study of a community of student designers where distributed creativity was theorized to occur. I conclude with implications from this research for the design and research of distributed creativity within student communities.

## Distributed Cognition

Distributed Cognition theories emerged in the 1990s as companions to situated cognition theories, and in fact bear many similar definitions, constructs, and theoretical foundations (Moore & Rocklin, 1998). Researchers developed this theory as an alternative to traditional information processing models of cognition, which often neglected the impact of social variables, by integrating ideas from anthropology, social psychology, sociology and the Russian cultural-historical school of psychology (Cole & Engestrom, 1993). Salomon (1993) believed there were three reasons for the development of distributed cognition theories: 1) the increasingly important role of technology for intellectual tasks—something that has grown exponentially true since Salomon's original essay, 2) the re-emphasis on Russian cultural-historical theories, and 3) dissatisfaction with the limitations from conceiving of cognition as bounded within individuals.

In the past two decades of theoretical and research-based work on distributed cognition theories, different conceptualizations have emerged. For example, Hutchins (1995) wrote that socially distributed cognition has a parallelism that can't be found in individual processing because multiple complex tasks can be simultaneously completed within a system, but not within an individual. Second, tools (including language) always mediate communication between persons in a system, and this creates a problem of the "bandwidth of communication" (p. 284) between members of the system. Hutchins gave the example of a complex navigation system that had some cognitive processes held interpsychologically between the system members that could "never be internalized by a single individual" (P. 284). Hutchins believed that "all divisions of labor, whether the labor is physical or cognitive in nature, require distributed cognition in order to coordinate the activities of the participants" (p. 176). Nardi (1996) added the understanding of "functional systems" to redirect analysis to the systems level, or the level of individuals and artifacts, and the coordination between these two. He explained that distributed cognition focused on structure, or representations, both inside and outside a person's individual cognition. Another way of conceptualizing distributed cognition might be through the kinds of interactions used to distribute the processing. For example, Hwang, Hsu, Tretiakov, Chou, and Lee (2009) explored the roles of interaction (productive collaboration), overaction (social, off-task, but necessary conversation between people), and intra-action (distributing thinking within one person through note taking, tool use, etc.). They found that intra-action had the strongest relationship with learning outcomes, suggesting the particularly important role that tools and non-human systems can play in cognition.

Moore and Rocklin (1998) summarized distributed cognition as representing two different frameworks. The first, "individual-plus," conceives of cognition as primarily residing within individuals but influenced by, and occasionally distributed among, the interacting social and artifact (tools) systems. Moore & Rocklin explained that this perspective conceptualizes cognitions as "divided among an individual, objects, and other people" (p. 107).

A second framework described by Moore and Rocklin (1998) is that of the "social-only" perspective, which conceives of cognition as so thoroughly distributed that one cannot speak of cognition as an individual process because it resides within the group. This brings a realization that "not only do social and other situational factors have an impact on cognitions that occur in one's mind, but that the social processes *themselves* should be considered cognitions (Greenberg & Dicketman, 2002, p. 19, emphasis in original). Cole and Engeström (1993) expanded this framework with a description of Activity Theory, emphasizing cognition and problem solving as occurring through the interactions between individuals, the environment, and the tools or artifacts (conceptual and physical) of the environment. From this view, the "natural unit of analysis for the study of human behavior is activity systems" (p. 9)—or in other words, the study of cognition within social contexts. Thus, when discussing distributed cognition, there are various interpretations of how distributed the cognitive processes must be, leaving room for additional research and clarification.

## Theoretical Application to Distributed Creativity

Traditionally, the psychological study of creativity focused on individual perspectives, taking its cue from information processing cognition models (Mandler, 1995). A trend towards representing individual creativity with cognitive terms and models has also gained support (Smith, Ward, & Finke, 1995; Ward, Smith, & Finke, 1999). However, beginning in the late 1980s and 1990s, some creativity researchers started describing creativity in ways similar to the "individual-plus" model of distributed cognition. For example, Csikszentmihalyi (1999) developed a systems model of creativity, explaining that individual creativity is influenced and defined by the system in which it resides. Other researchers explored the role of the social climate within an organization that might enable innovation (Anderson and West, 1996; Amabile et. al, 1996), or how knowledge is structured and managed within creative enterprises (McAdam, 2004).

Sawyer and DeZutter (2009) have recently described distributed creativity differently, in a way comparable to the *social cognition* branch of distributed cognition. They argue that "creativity is embedded in social groups" and "significant creations are almost always the result of complex collaborations" (p. 81). They defined distributed creativity as pertaining to collaborating groups that collectively produce a creative product, in either predictable and constrained, or unpredictable and unconstrained, environments. They believe that collaborative emergence, or the unpredictable and unexpected emergence of distributed creativity, occurs when activities have unpredictable outcomes; interdependency within the group such that a person's actions are influenced and constrained by the actions of others; and collaboration (equal member contribution).

A possible framework representing Sawyer and DeZutter's conception of distributed creativity within adult learning and working communities could be that of Communities of Innovation (West, 2009). This framework was developed by combining principles drawn from theoretical and research-based discussions in psychology, social learning theory, and organizational development. It proposes that innovation emerges in communities that have the following characteristics:

- Dynamic expertise, characterized by "continuous efforts to surpass one's earlier achievements and work at the edge of one's competence" (Hakkarainen, Palonen, Paavola, & Lehtinen, 2004, p. 243)
- Group flow, which Sawyer (2008) described as including: 1) a shared goal, 2) close or deep listening to each other, 3) complete concentration, 4) being in control of the group's actions and environment, 5) blending of individual egos, 6) equal participation, 7) members' familiarity with each other, 8) constant communication, 9) elaboration of each others' ideas, and 10) frequent failure (and learning from failure).
- Entrepreneurship and ownership. Innovative communities need to develop the unique type of environment that allows enough structure to keep the community together and focused, but enough flexibility to allow individual members to take ownership over their own projects and ideas (Coakes & Smith, 2007; McFadzean, O'Loughlin, & Shaw, 2005)

- Inquiry. Hakkarainen et al. (2004) found that "all models of innovative knowledge communities . . . highlight the role of problems and questions that guide the process of knowledge creation" (p. 197).
- Group reflectiveness. Hakkarainen et al. (2004) argued that both interpersonal and intrapersonal reflection was important and others incorporated similar ideas in their models (Bielaczyc & Collins, 2006; Engeström, 1999; Sawyer, 2008).
- Diversity. Justesen (2004) described diversity in techne and cognition as so critical that she described it as "innoversity." Others have echoed these sentiments (Bielaczyc & Collins, 2006).
- New community boundaries, visions, and goals. In COIs, it is more likely that members network with persons both within and outside the community (Sawhney & Prandelli, 2000) and are less controlled by management and typical beauracracy (Benton & Giavagnoli, 2006).
- "Hacker"-like motivation. Himanen (2001) and Raymond (2003) explained that hackers care deeply about their work and quality, becoming so intensely motivated by their projects that it becomes almost playful or joyful. Hacker-like motivation can be found in areas outside of computer programming where innovation emerges.

These principles, drawn from these and other sources, were organized into the Communities of Innovation framework, and formed the basis for this study, which was an initial exploration of how robust this framework could be when applied to a higher educational setting.

## Research Questions and Methods

The purpose of my research agenda was to study the nature of distributed creativity from a Communities of Innovation perspective, within student design communities in the context of higher education. Specifically, my research questions for this study were:

1. Do elements of a community of innovation emerge among members of a graduate instructional design studio?
2. If so, how do members of this community describe those elements? If not, what do members report might have impeded the development of a COI in this setting?

### Research Design and Participants

This study combined phenomenological interviewing (Seidman 2006) and Critical Incident Technique (Flanagan, 1952) to study the emerging distributed creativity during one semester of a design Studio for graduate-level instructional designers at a large, Southern university. This Studio consisted of three courses: A beginning course focused on gaining expertise utilizing design technologies; a second course involving an individual design project, and a third course involving a larger, group design project. There were high levels of collaboration among students within each course, as well as students between the different courses as newer students were required to serve as assistants for the more experienced students' projects, and more experienced students were required to mentor newer students in their projects. All three courses met together for general instruction and discussion of design principles each week. From this setting, four students (Jamie, Robin, Boyd, and Lori), representing all three Studio courses, were selected as case studies based on their background and inclination towards collaborating in their creative works.

### Data Collection Methods

A combination of methods were used to explore the nature of distributed creativity within this Studio community. This research study followed Seidman's (2006) strategy for phenomenological interviewing, except for one modification. Seidman outlines a three-interview process. The first interview is designed to understand the participant's background relevant to the experience at hand. I conducted this interview at the beginning of the semester about the participants' previous engagements in creative and collaborative school activities. Seidman then recommends a second interview for specific details about the experience itself, and a third interview where the participant and the researcher co-interpret the experience to understand its significance and meaning.

In lieu of Seidman's (2006) second interview, and in order to better understand the specific details of how ideas developed through group collaboration, each participant recorded a weekly 5- to 10-minute voice memo detailing the "critical incidents" (Flanagan, 1952) of their group design activities. In other words, they explained 1) what happened that week that was significant to their project, 2) who was involved in the incident, and 3) why this incident positively or negatively impacted their project. This method enabled the students to provide more thorough details into their weekly group collaborations than they would have been able to remember in a single interview. This information was triangulated with an analysis of the students' design experiences as reported in their required weekly design journals. I also observed first-hand many of the participants' interactions with their peers.

I then concluded the semester by conducting the third of Seidman's interviews and co-interpreting with the participants how their group actions enabled or frustrated the creative process. In addition to these data sources, I interviewed the Studio instructors in order to understand their rationale, goals, and perspectives on the Studio setting.

## Data Analysis Methods and Rigor Guidelines

Data were analyzed using constant comparison coding techniques for forming categories and theories derived from categories. This analytic method is similar to the form of analysis suggested by Flanagan (1952) for Critical Incident data, which I collected via the student voice memos. Initial categories were derived from the COI model (group flow, hacker ethic, entrepreneurship, etc.). Additional categories were created from the analysis process to describe emerging significant events or patterns in the participants' experiences. Trustworthiness was developed through member checking case study reports with the participants, asking independent coders to analyze uncoded portions of the data to confirm emerging themes, peer review of the theoretical and methodological frameworks, and triangulation of methods and data sources.

## Findings

Findings included the emergence of several elements of the Community of Innovation framework in the experiences of the participants (flow, hacker ethic, and entrepreneurship). In addition, new themes emerged representing additional ideas that could be added to the framework (perspectives on collaboration and mentoring, interactive idea generation, sense of community, learning through design criticism, and idea prototyping). Finally, some COI elements were not found in this study (see Figure 1).

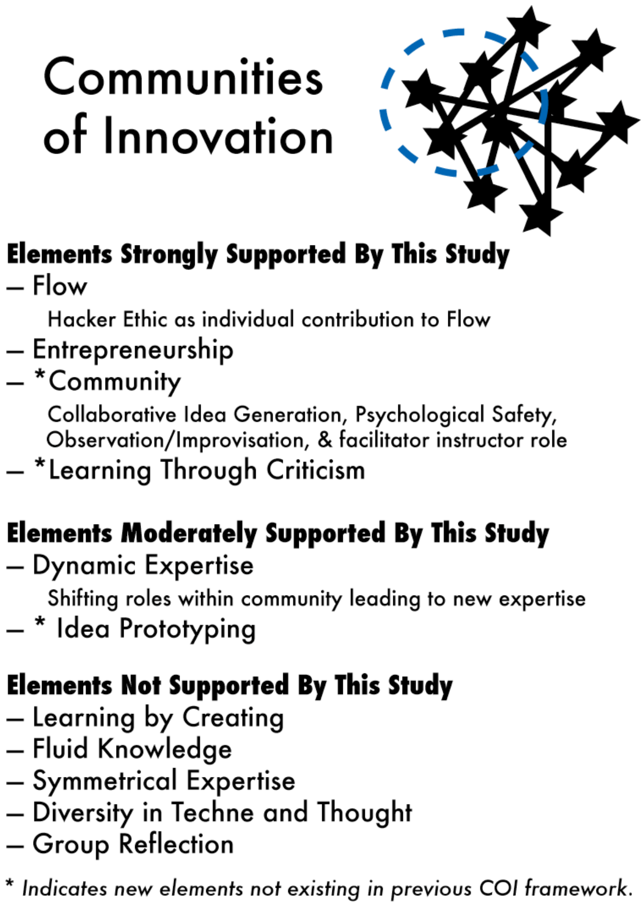



Figure 1. Revised Communities of Innovation framework

### Existing COI Elements Supported by the Data

#### Flow

While Group Flow (Sawyer, 2008) was theorized to be specifically important to COIs, it was difficult to detect Group Flow in this study (12 coded statements). This could be due to the lack of group flow or to the difficulty of capturing it within the questions and self-reports used in this study. However, individual flow (Csikszentmihályi, 1990) emerged as a significant theme (30 statements, found in all four case studies' experiences), and was coded as situations when participants reported being completely engaged to the point of losing consciousness of their surroundings and time, in part because they felt competent to complete the task, understood the bounds and rules of the activity, and found it personally enjoyable. While the flow experience was mostly described as individual, it appeared to be fostered by community characteristics such as project ownership; student agency; and a fertile, creative environment.

#### Hacker ethic

Hacker ethic was coded when statements described the participants' work as interesting, playful, or completed because of a desire for quality or satisfaction rather than a grade, or involving high levels of enthusiasm. From the literature (Himanen, 2001; Raymond, 2003) and this research, I interpreted flow as an experience that happens to people, whereas a hacker ethic is something innate that a person brings to an experience. Data in this study supported this distinction, but participants often indicated that both existed simultaneously. All four participants described having a hacker ethic for learning the skills needed to complete high-quality projects, although Jamie least so (6 coded statements compared with an average of 22 statements for each of the other three).

## Entrepreneurship and autonomy

Entrepreneurship/autonomy was coded when participants discussed the ability to create their own projects, define their own goals, or take ownership in their work. All participants indicated that Studio enabled them to become innovative by allowing autonomy in selecting and designing projects. Boyd remarked, "It [Studio] lets you have your own goals not compared to somebody else" and "they really give you free reign" over tools, learning design/theory, etc. He continued, "Really the only thing that can impede you . . . would just be your own limitations."

## Collaboration and mentoring

"Collaboration," defined as repeated interactions focused on achieving a goal such as developing a project component or learning new skills, was evident in 173 comments. All four participants reported that collaboration was crucial to developing their projects, although they defined collaboration differently and benefited from different kinds of collaborative relationships. The participants often indicated a desire for even more collaboration (32 statements). Interestingly, despite all four participants being very comfortable with Internet technologies and two of them living far away from the university, only Lori indicated collaborating with friends on the Internet, while the others strongly preferred face-to-face collaborations.

Some of the collaborations reported by the participants were minor interactions with other community members that either pulled the participants away from or reinforced a particular decision, or gave emotional support for a chosen action. These small nudges in a particular direction were coded 34 times. Sometimes, however, the participants indicated more dedicated, consistent, and one-on-one collaboration that was coded as mentoring (30 statements). All participants reported some degree of mentoring, usually to support their technical skill development.

## Interactive idea generation

In this study, we flagged every instance where participants mentioned a new idea, and coded these ideas as having originated from the participants themselves, through interactions with others, or from materials such as textbooks or tutorials. In general, participants reported mostly receiving ideas through interactions with others, especially other Studio members (134 coded statements, see Figure 2), as well as from connections outside of Studio (37 statements). Participants also drew ideas from assigned textbooks (8 statements), and from searching on the Internet (41 statements). Mostly they reported their ideas were generated interactively with others. Some of these ideas were related to technical issues and learning new technologies. Most ideas were related to minor design changes, usability issues, and aesthetic improvements. Robin eventually changed the entire template for her project because Boyd, in a desk critique, questioned the viability of her previous design. In return, Robin and others offered ideas to Boyd about adding interactive elements to his Web site and improving his font and color choices. Boyd noted*,* I received some really great feedback, . . . which led directly to changes

**Where Participants Reported Receiving Ideas**

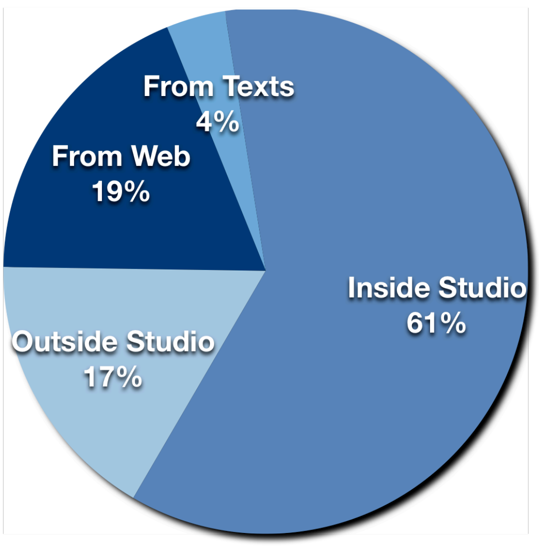



Figure 2. Where participants reported receiving their ideas.

## Sense of community

Sense of community was coded a total of 50 times when participants indicated being emotionally or psychologically connected with, trusting, receiving support and encouragement from, and feeling friendly with their Studio peers. [Note: Although collaboration could be another indication of the strength of a community, I coded collaborative events separately to allow specific analysis of those interactions.]

## Learning through critiquing

Learning through critiquing was coded when participants indicated learning or gaining insights from the peer feedback process or from evaluating other designs, and it was coded 39 times. For example, Lori mentioned that she and a friend sat in on each others' prototyping meetings with their instructor. During the discussion of her friend's prototype, Lori contributed advice that caused her to reflect on her own project: "In just some of the things I suggested to her I was like wait a minute, I could be doing that for my project." She further observed, "so much more happens when you can actually sit and talk about your project."

#### Idea prototyping

Participants discussed the importance of prototyping ideas to facilitate idea generation through observation and improvisation. However, the evidence for including this element in the COI framework is still tentative (13 coded statements in this study). This concept is supported by models of rapid prototyping, which is an approach to design that emphasizes a "rapid, iterative series of tryout and revision cycles . . . until an acceptable version is created" (Baek, Cagiltay, Boling & Frick, 2008, p. 660). Often, rapid prototyping involves users in testing the product, but this study indicated that it was also necessary for engaging members of a COI in developing innovative ideas. Participants suggested that prototyping might be most influential when it begins early in the design process and when sufficient to facilitate one-on-one or small-group discussions about the prototype.

### Challenges to Implementing a COI

From this study, several challenges to implementing a Community of Innovation in an educational setting were evident. First, there was a lack of time for completing the tasks sufficiently, causing participants to often focus solely on completing tasks instead of considering the most innovative or effective way to produce their projects. Another impediment for the students was their lack of technology skills, particularly for Boyd and Lori who had innovative ideas but could not develop them due to limited technical skills. Thus, while Hakkarainen, et al. (2004) noted the importance of dynamic expertise that is adaptable to changing problems, there appears to be a need also for domain-specific expertise as a prerequisite to innovative collaboration and improvisation.

Finally, participants reported receiving ideas, social support, and feedback from peers, but this support was usually superficial unless it came from a member of their close peer group. The clearest example was Lori, who described working closely with a dedicated and skilled mentor, but reported almost no collaborations with anyone else. Robin and Boyd did not have dedicated mentors but instead formed a group with Studio friends that provided quality feedback and support. Like Lori, however, they collaborated little with anyone else. Jamie worked closely with her team members, but reported little interactions outside of her team. Thus, COI support and collaboration may impact innovation only among members of local, helpful peer groups within the community. If so, connecting community members with "innovation champions" (Coakes & Smith, 2007) or developing expert networks (Hakkarainen et al., 2004) may prove especially important.

## Conclusions: Reexamining the Formative COI Framework

In this study, I employed a formative Communities of Innovation framework, an adaptation of theories about communities of learning/practice and creativity research, to describe the innovative potential of adult groups. Not all of the theorized COI elements were evident in the data. Findings included evidence for some aspects of the proposed COI model (flow and hacker ethic, entrepreneurship, collaboration and mentoring, interactive idea generation, sense of community, and learning through design criticism), moderate support for others (dynamic expertise and idea prototyping), and no evidence supporting other proposed components from the previous iteration of the model (West, 2009—developing adaptable knowledge and expertise, symmetrical expertise within the community, community reflection, shifting interpersonal roles, or benefiting from cultural/ educational/skill/other diversity). There was not evidence that these latter components are not important to a COI, only that they weren't evident in this study.

Based on these findings, the COI framework was tentatively revised to differentiate original elements that were strongly and weakly supported by data in this study, new elements supported by the data, and those original elements that were not supported by the evidence in this research (see Figure 1). This distinction is helpful for designing and researching COIs, as it creates priorities for emphasizing specific elements in a given community.

## Implications for Future Research

As Greeno (1997) explained, we need to seek to understand "which combinations and sequences of learning activities will prepare students best for the kinds of participation in social practices that we value most" (p. 9). Because graduates enter a workforce that is increasingly demanding creativity as the currency for success (Banahan & Playfoot, 2004), it is important to continue researching how we can foster effective distributed creativity in higher education in order to best replicate the kinds of social practices and activities graduates will engage in when they fully enter society. The conceptualization of COIs described in this paper is formative. Support for some components was apparent in this study, but several unanswered questions for future research remain, including:

1. What is the nature of group flow and how can it be developed within a community?

Flow was common across participants' experiences, but was usually manifested as individual rather than group flow, which may have been due to methodological limitations. Future research is warranted to delineate the differences between individual and group flow, articulate the nature of individual and group flow, and examine what influences group flow. To address these questions, conversation analysis—a methodology designed to rigorously capture routine, everyday activities occurring in naturalistic settings in a manner that is reproducible and defensible (Sawyer & DeZutter, 2009)—might be used.

2. How do COI designers balance structure and scaffolding with autonomy?

In this study, autonomy and entrepreneurship were key characteristics of COIs. Yet, this freedom created challenges for participants as they struggled to identify the vision for their projects. Future research is needed to understand how to balance structure and scaffolding, especially for novices, and the autonomy needed to promote innovation. Researchers might employ quasi-experimental studies with control and experimental groups to account for varying levels of scaffolding and structure. Results could be compared according to expert judgments of the innovativeness of the final products, or by utilizing a measure of creative potential (Kim, 2007) or divergent thinking (Runco, 1993). Qualitative methods could explore the nature of the scaffolding found to be most effective, and how participants perceived, experienced, and benefited from this scaffolding.

3. What is the nature of community within a COI and how does this compare with other communities?

Research is needed to articulate the nature of the community within a COI and how this differs from other kinds of communities, such as learning communities and communities of practice. These questions could be studied via social network analysis in order to quantify the social capital of relationships making up COIs.

4. How is knowledge and expertise acquired in a COI?

This study provided tentative findings related to how innovation develops through the peer critiquing process and that how dynamic expertise influences innovation. However, research is needed to verify these findings and extend our understanding of these principles. Case study methods could prove valuable for documenting how dynamic expertise is developed, relying on a combination of critical incident recall and close researcher observation with a small participant sample. Video analysis may also be helpful in capturing the nuances of expertise development. Conversation analysis could again be useful in microanalyzing the discourse.

5. What is the value of COIs? Do they produce more innovative ideas or products?

A significant, and largely unanswered, question concerns whether COIs stimulate more innovation than other social structures. Researchers could use historical approaches by first identifying major innovative ideas and working backwards to analyze archival data concerning the social structure surrounding the innovation. Another approach would be for experts to review the products generated by a COI and other kinds of communities to develop a reliable instrument for analyzing the innovative potential of group ideas.

## References


Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal, 39*(5), 1154-1184.

Anderson, N., & West, M. A. (1996). The Team Climate Inventory: Development of the TCI and its Applications in Teambuilding for Innovativeness. *European Journal of Work & Organizational Psychology, 5*(1), 53.

Banahan, E., & Playfoot, J. (2004). Socio-organisational challenges in the creative economy. In L. M. Camarinha-Matos & H. Afsarmanesh (Eds.), *Collaborative networked organizations: A research agenda for emerging business models*. Boston: Kluwer Academic Publishers.

Baek, E.-O., Cagiltay, K., Boling, E. and Frick, T. (2008). User-centered design and development. In J. M. Spector, M. D. Merrill, J.J. van Merrienboer & M. F. Driscoll (Eds). *Handbook of research on educational communications and technology* (3rd ed.). 659-670.

Benton, S., & Giovagnoli, M. (2006). *The wisdom network: An 8-step process for identifying, sharing, and leveraging individual expertise*. New York: American Management Association.

Bielaczyc, K., Collins, A., O'Donnell, A. M., Hmelo-Silver, C. E., & Erkens, G. (2006). Fostering Knowledge-Creating Communities. In *Collaborative learning, reasoning, and technology.* (pp. 37-60): Lawrence Erlbaum Associates Publishers.

Coakes, E. & Smith. P. (2007). Developing communities of innovation by identifying innovation champions. *Learning Organization, 14*(1), 74-85.

Cole, & Engeström (1993). A cultural-historical approach to distributed cognition. In *Distributed cognitions: Psychology and educational considerations.* (pp. 1-46). Cambridge: Cambridge University Press.

Csikszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. New York: HarperCollins.

Csikszentmihalyi, M. C. (1999). Implications of a systems perspective for the study of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity.* (pp. 313-38). Cambridge: Cambridge University Press

Engestrom, Y. (1999). Innovative learning in work teams: Analyzing cycles of knowledge creation in practice. In Y. Engestrom, R. Miettinen & R.-L. Punamaki (Eds.), *Perspectives on activity theory* (pp. 377-404). Cambridge: Cambridge University Press.

Flanagan, J. C. (1952). *The critical incident technique in the study of individuals*. Washington D.C.: American Council on Education.

Greenberg, 2000 is what is cited.
Greenberg, J. D., & Dicketman, G. J. (2002). Distributed cognition: A foundation for performance support. *The ASTD E-Learning Handbook*, 303.
Greeno, J. G. (1997). Response: On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5-17.
Hakkarainen, K., Palonen, T., Paavola, S., & Lehtinen, E. (2004). *Communities of networked expertise: Professional and educational perspectives*. Amsterdam: Elsevier.
Himanen, P. (2001). *The hacker ethic: A radical approach to the philosophy of business*. New York: Random House.
Hutchins (1995). *Cognition in the wild.* Cambridge: MIT Press.
Hwang, W., Hsu, J., Tretiakov, A., Chou, H., & Lee, C. (2009). Intra-Action, Interaction and Outeraction in Blended Learning Environments. *Educational Technology & Society*, *12*(2), 222-239.
Justesen, S. (2004). Innoversity in communities of practice. In P. M. Hildreth & C. Kimble (Eds.), *Knowledge networks: Innovation through communities of practice* (pp. 79-95). Hershey, PA: Idea Group Publishing.
Kim, K.-H. (2007). The two Torrance creativity tests: The Torrance Tests of Creative Thinking and thinking creatively in action and movement. In A.-G. Tan (Ed.), *Creativity: A handbook for teachers*: World Scientific.
Mandler, G. (1995). Origins and consequences of novelty. In S. M. Smith, T. B. Ward & R. A. Finke (Eds.), *The creative cognition approach* (pp. 9-26). Cambridge: MIT Press.
McAdam, R. (2004). Knowledge creation and idea generation: A critical quality perspective. *Technovation*, *24*(9), 697-705.
McFadzean, E., O'Loughlin, A., & Shaw, E. (2005). Corporate entrepreneurship and innovation part 1: the missing link. *European Journal of Innovation Management, 8*(3), 350 - 372.
Moore, J. L., & Rocklin, T. R. (1998). The distribution of distributed cognition: Multiple interpretations and uses. *Educational Psychology Review*, *10*(1), 97-113.Nardi (1996). *Context and consciousness: Activity theory and human-computer interaction.* Cambridge: MIT Press.
Nardi, Bonnie A. *Context and Consciousness: Activity Theory and Human-Computer Interaction.* Cambridge: MIT Press, 1996.
Raymond, E. S. (2003). *The art of Unix programming*. Retrieved March 21, 2008, from http://www.faqs.org/docs/artu/ch01s09.html.
Runco, M. A. (1993). Divergent Thinking, Creativity, and Giftedness. *Gifted Child Quarterly, 37*(1), 16-22. doi: 10.1177/001698629303700103.
Salomon, G. (1993). Editor's introduction. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations*. New York: Cambridge University Press.
Sawhney, M., & Prandelli, E. (2000). Communities of creation: Managing distributed innovation in turbulent markets. *California Management Review, 42*(4), 24-54.
Sawyer, R. K. (2008). *Group genius: The creative power of collaboration*. Cambridge: Perseus Books Group.
Sawyer, R. K., & DeZutter, S. (2009). Distributed creativity: How collective creations emerge from collaboration. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(2), 81-92.
Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed., p. 162). New York: Teachers College Press.
Smith, S. M., Ward, T. B., & Finke, R. A. (1995). *The Creative Cognition Approach.* Cambridge: MIT Press..
Ward, T. B., Smith, S. M., & Finke, R. A. (1999). Creative cognition. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 189-212). Cambridge: Cambridge University Press.
West, R. E. (2009). What is shared? A framework for understanding shared innovation within communities. *Educational Technology Research and Development*, *57*(3), 315-332. doi: 10.1007/s11423-008-9107-4.

# Students' Meaning Making in a Mobile Assisted Chinese Idiom Learning Environment

Lung-Hsiang Wong, Learning Sciences Laboratory, National Institute of Education, 1, Nanyang Walk, Singapore 637616, lunghsiang.wong@nie.edu.sg
Chee-Kuen Chin, Chee-Lay Tan, May Liu, & Cheng Gong, Singapore Centre for Chinese Language, 287, Ghim Moh Road, Singapore 269623, {cheekuen.chin, cheelay.tan, may.liu, cheng.gong}@sccl.sg

**Abstract:** In recent years, we witness the rise of communicative and contextualized language learning approaches that is concomitant with developments of Mobile Assisted Language Learning (MALL). In this paper, we present a pilot study in MALL that emphasizes "creative learner outputs" and contextualized meaning making. In learning Chinese idioms, students used smartphones on a one-device-per-person basis to capture photos of the real-life contexts pertaining to the idioms, and to construct sentences with them. Subsequently, in-class or online discussions took place, which would enhance the students' understanding in the proper usage of the idioms. The learning design is grounded in the seamless learning model that encompass in-class formal learning and out-of-class informal setting, and personal and social learning spaces. The students' ongoing, open-ended, personal-to-social meaning making process and artifacts have indeed shown some indicators of "seamless language learning" that has the potential of transforming language learning into an authentic learning experience.

## Introduction

A closed system may settle into an undifferentiated state, while an open one is more likely to become animated, and sometimes highly coordinated (Kauffman, 1995). Likewise, in language learning, "closed," mechanical exercises restrict information to only "correct" answers that are unlikely to remain in permanent memory (Stevick, 1996), whereas meaningful and communicative activities build on the classroom to learners' wider knowledge—an almost inexhaustible resource (Bolte & Herrlitz, 1986).

This paper reports on a pilot study of Mobile Assisted Language Learning (MALL) in Nan Chiau Primary School in Singapore. In the study, we facilitated a 5$^{th}$ Grade class to study and apply 29 common Chinese idioms, within the context of learning Chinese as a Second Language (L2). Apart from in-class idiom lessons with contextualized learning activities, the students were assigned with a smartphone each which they kept and accessed 24x7 throughout the nine-week period of the study. They were instructed to carry out photoblogging-like activities by using their smartphones to take photos in their daily lives, and then make sentences with the learned idioms, and subsequently to post them onto a wiki space for sharing and peer review.

The MALL design emphasizes students' proactive association of the contexts that they encounter in the real world with the learned Chinese idioms (a special form of vocabulary). In this paper, we focus on studying the students' individual-to-social learning process, which could be attributed to the process of multimodal, student artifact-focused, ongoing, open-ended meaning making in vocabulary learning. The students' learning processes and outcomes suggest a compelling direction for MALL – seamless language learning.

## Literature Review

### Constructivist Approach in Language Learning

In recent decades, we witness a paradigm shift of language learning theories from behaviorism and structurism to a communicative approach (Salaberry, 1996). The emphasis of the communicative approach is on real-world communication in environment and situation that language learners might actually experience (Hoopingarner, 2009). Under the emerging paradigm, learning is seen not as a passive activity that requires learners to accept pre-packaged information, but as an active process by which learners create their own understanding. This approach meshes well with the constructivist theory of learning (Jonassen, 1991).

Language learning involves active mental activity such as interaction and hypothesis testing (Gass, 1997) and productive output (Swain, 1985). In addition, prior studies in the L2 classroom have brought our attention to the importance of the negotiation of meaning, also known as social meaning making, in L2 development (e.g., Long, 1985). More sophisticated meaning making involves the mutual transformation in the actions, and perhaps the thoughts and moods, of two partners (Locke & Bogin, 2006). However, the social context has been undervalued as an arena for collaborative L2 learning. Where meaning appears fixed, immutable, to be sent and received, what is lost is the collaborative nature of meaning making (Savignon, 1991).

### Vocabulary (Idiom) Learning

As a main component of language learning, vocabulary learning is often delivered in conventional ways, such as providing abstract definitions and sentences taken out of the context of normal use. Such strategies might well lead to simplistic conception of vocabulary (Miller & Gildea, 1987), as they may deprive the students of pragmatic cues and render the process of meaning making harder (Kumaravadivelu, 1994).

Vocabulary can be classified into two categories, context-free and context-dependent vocabulary. Context-free vocabularies, such as nouns, are those which can stand by themselves without depending on sentence contexts (Elliot & Zhang, 1998). Context-dependent vocabularies could either be prepositions, which are not rich in meaning and are better learnt in sentence as their meanings depend on the presence of other words (ibid), or compound vocabulary such as idioms whose complex nature may result in context-dependent appropriateness of their usage (Deng, 2001) – there are many possible real-life contexts where such vocabulary could suitably (or unsuitably but often mistakenly) be used. As contemporary learning theorists argue that language teachers should create the right conditions for students to 'uncover' grammar (e.g., Thornbury, 2001) through active meaning making (Bourke, 2005), we envisage a similar principle for vocabulary learning.

Over the years, vocabulary learning theorists have advocated productive (apart from receptive) learning (Webb, 2005), inductive meaning making (Lewis, 1997), and contextual learning (Sanaoui, 1995), among others. Furthermore, Nation (2001) proposes three psychological processes for successful vocabulary learning: noticing (a word is highlighted as being salient text input), retrieving (repeat encountering of the word) and creative/generative (a previously encountered word is met or used in a slightly different context). The three-stage model stresses the importance of the coupling of receptive and productive learning, and the learners' generative usage of the vocabulary in alternative contexts. All these point to the trend of emphasizing students' self-construction of understanding in vocabulary usage, and this is done most likely through authentic learning.

### Mobile Technology for Vocabulary Learning

As authentic learning (i.e., learning activities that are framed around real-life contexts) comes into the picture of language learning (Widdowson, 1978), MALL becomes a viable solution to the blending of the language learners' learning environment into their real-life contexts. Prior research on mobile learning has shown that the mobility and connectivity of the devices enable students to become an active participant, not a passive receiver, in learning activities (Looi *et al.*, 2009). The recent development of MALL demonstrates a similar tendency.

According to a survey by Kukulska-Hulme & Shield (2007), prior MALL studies could be divided into two categories: content-based (formal learning-oriented; developing learning materials for mobile access) and design-oriented (informal learning-oriented, out-of-class, often authentic and/or social learning activities). Most of the MALL studies in the context of vocabulary learning have conformed to content-based design, whose systems work in the way of pushing of relatively static learning materials or quizzes to the device learning (e.g., Thornton & Houser, 2002; Levy & Kennedy, 2005), which could be attributed to behaviorist learning despite having the advantage of anytime, anywhere. Ogata, Akamatsu & Yano's (2004) design moves one big step by incorporating context-aware technology, as their system enables the learners to use their handhelds to detect the RFID-tagged objects and retrieve their names (vocabulary) and explanations. All the above-stated designs are however still restricted to receptive learning and are usually (not always) only suitable for learning context-free vocabulary. Learning context-dependent vocabulary with the above designs may result in learners' understanding of the general definitions but not necessary the proper usage of them.

### Seamless Learning and Language Learning

Recognizing both the importance and the limitation of formal, in-class language learning, language learning theorists have been advocating the integration of formal and informal language learning (which could be dated back to the pre-MALL days, e.g., Titone, 1969). MALL becomes a solution to the blending of the language learners' learning environment into their real-life contexts. The handhelds which could function as a personal "learning hub" creates the potential for a new wave of evolution of technology-enhanced learning. This evolution is characterized by "seamless learning spaces" and marked by continuity of the learning experience across different environments, and emerging from the availability of one device or more per student ("one-to-one") for 24x7 access (Chan *et al.*, 2006). In particular, the integration of individual and social learning could be further enhanced by blending mobile and Web 2.0 technologies to bring to the students the situated mobile learning experiences that take into account both the students' everyday tasks and socio-constructivism (Winters, 2007). Such integration can be expected to balance and bring out the best of both individual and social learning.

The integration of technology-assisted informal learning into the students' lives does not imply the elimination of school-based formal learning. Instead, it becomes more challenging for teachers to inspire students' learning in both formal and informal settings by modeling the seamless learning process (Zhang *et al.*, 2009). In this regard, we are keen to tap on the seamless learning framework to carry out the modern language learning pedagogy for the e-generation. Such framework favors contextualized learning, the integration of formal and informal learning, and, more importantly, the integration of receptive and productive learning.

## Study Description

Our one-and-a-half-year study, "Move, Idioms!" (成语，动起来！), consists of a pilot study and a full-scale intervention, in July-September 2009 and in January-October 2010 respectively. This paper focuses on the design, implementation, and our analysis on the students' artifacts of the completed pilot study.

In facilitating the pilot study, we designed a learning process to engage students in ongoing Chinese idiom learning and writing (sentence making) activities. A class of 40 Primary 5 students participated in the study. Each of them was assigned a HTC TyTN II smartphone with built-in digital camera, Wi-Fi access, internet browser and Chinese text input. Furthermore, we used PBWorks (http://www.pbworks.com/) to create the wiki space for photo/sentence sharing and peer reviews. Apart from standard wiki features such as multi-user content editor and page history, an online forum-style comment tool is incorporated on each wiki page. In addition, mobile-optimized animations that depict the meanings of idioms can be accessed by students anytime, anywhere. The animations are sponsored by our research partner, a digital content developer based in Taiwan.

Figure 1 depicts the process of our learning design. The four-activity process is iterative and encompasses formal and informal learning spaces, individual and social learning spaces, receptive and productive activities, and the use of both mobile and Web 2.0 technologies (that is, learning takes place in both the physical world and the cyberspace). The processes of the four activities are described below:

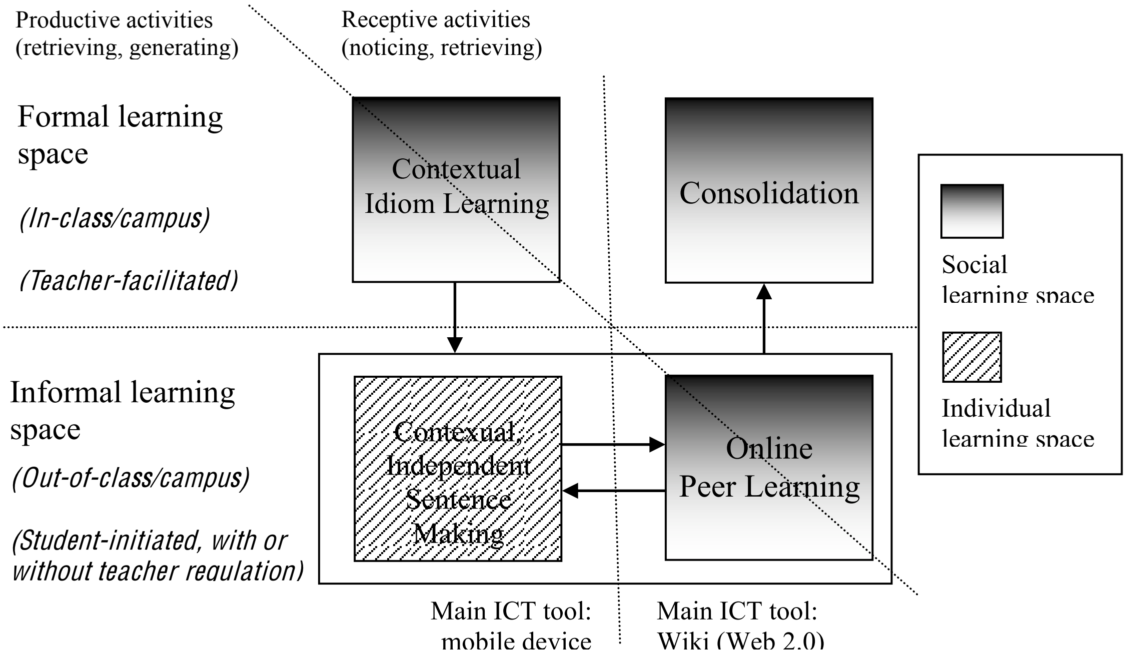



Figure 1. The mobile-assisted idiom learning process.

Activity 1 – In-class contextual idiom learning: The classroom/in-campus activities, to be designed in the form of lesson plans, are conducted with the aim of carrying out the "noticing" and "retrieving" processes in Nation's (2001) framework as well as preparing students to engage in subsequent out-of-school activities ("generating"). During each lesson, a new set of "idiomatic animations" is shown to the class (also accessible by the students after school). The teacher then conduct contextualized learning activities, such as creating linguistic contexts, for instance by providing context-rich images, inviting students to discuss about relevant idioms, and even "mobilizing" the students to take photos in the campus to illustrate the idioms and upload them to the Web.

Activity 2 – Out-of-class, contextual, independent sentence making: students carry the phones assigned to them 24x7. Apart from watching the animations repeatedly, students are instructed to identify or create contexts in their daily lives which could be associated with any learned idiom. They then take photos, make sentences that contain the idioms to describe the photos, and post them onto a class wiki space. In the wiki space, we create one wiki page for each idiom covered in the class for students to post their photos/sentences. This offers convenience for comparing various student-identified contexts pertaining to the same idiom.

Activity 3 – Out-of-class, online peer learning: students are encouraged to perform peer reviews on the wiki by commenting on (with the PBWorks comment tool), correcting or improving their peers' sentences (by making direct modifications on the sentences posted on the wiki pages with a different font color). During the pilot study, they carried out these activities with PC's at home, not the phones, due to technical constraint.

Activity 4 – In-class consolidation: Possible activities include class or small group discussions on selected sentences made by the students, or polls for "the most popular photo/sentence" on each "idiom page".

During the nine-week period of the pilot study, the teacher conducted five "idiom classes" (Activity 1) in the first five weeks with roughly one-week intervals. In the first three classes, the students were required to enact some of the idioms for peers to take photos within the classroom. In the last two classes, the students captured images to illustrate idioms within the campus. In between, the students were encouraged to carry out Activity 2 and 3. The teacher then facilitated Activity 4 in the seventh week. Students worked in groups of five, with each group being assigned an "idiom page" to discuss and identify erroneous uses of idioms with respect to the contexts in the photos or the sentences made, and to provide recommendation in correcting or improving the sentences. The students returned the phones to the school upon the completion of the study.

The data collected and analyzed for project evaluation consists of: (1) Pre- and post-quizzes; (2) Pre- and post-questionnaires; (3) Pre- and post-interviews with two high-, two medium- and two low-ability students (in terms of their academic performances in Chinese class) selected by the teacher; (4) Field notes taken during the in-class activities; (5) Student artifacts and online interactions. Due to space constraint and since we focus this paper on the analysis of the students' meaning making process, we will not go into the details of (1)-(4).

## Findings

### Constructivist Approach in Language Learning

Within the nine-week period, the 40 students contributed a total of 481 photo/sentence sets, revised (corrected or modified) sentences for 124 times, and posted 134 comments. However, the variation in the statistics of individual students' contributions was huge (mean = 12.0, SD = 25.9), as the top contributor posted 151 photo/sentence sets (or 31.9% of the entire class's postings) while 70.0% of her peers contributed less than 1 photo/sentence set per week in average. On the other hand, the students' participation levels were more consistent in offering sentence revisions (mean = 3.1, SD = 7.3) and comments (mean = 4.5, SD = 3.4).

Our observation and our interactions with the teacher and the students throughout the pilot study and our post-study interviews with six students had helped us to identify the challenges that had resulted in the widely varied levels of participation among the students, which are listed below:

- Affective issue: Many students showed great engagement during the in-class Activity 1 but when it came to the after school informal setting, they did not treat the smartphone as a learning tool but more a new toy. As what we have found out from the interviews and casual chats with the students, after enthusiastic exploration of the smartphone functionalities available in the initial period, many of the students had resorted to play online or installed games from time to time apart from carrying out the instructed activities. When being questioned about how they could and whether they had used their phones for other out-of-school learning activities, their common reply was: "no idea", except for a student who claimed that he had occasionally checked an online English dictionary with his phone in doing school assignments.
- Technical issue: Some students intermittently encountered and were frustrated by the technical problems in posting photos and sentences to the Web via their handhelds.
- Parental attitudes: Fearing of misplace or damage, most of the parents forbad the students to bring the phones out of their home except for bringing them to school, thus defeating the purpose of 24x7 seamless learning and severely narrowed the contexts that the students could associate the learned idioms with. Indeed, 83.2% of the contributed photos taken outside campus took place within the students' home.

As a learning design informed by the seamless learning framework, which implies the necessity of changing students' belief and, it is unrealistic to expect a significant breakthrough in the students within a nine-week period. Despite all that, after analyzing the student artifacts and their peer learning process, we found that they have shown great potential and promise, which we shall elaborate in the following sub-sections.

### Student Artifacts

While the earlier activities – teacher's explanation and students' viewing of the idiomatic animations – would be attributed as "transfer of encoded meanings" of the idioms, it is not the end of the students' learning journey and should instead serve as starting points for the students' subsequent "ongoing, open-ended meaning making" (Hedley, 1992) with the aim of rectifying, internalizing and enriching their understanding of the idioms.

From their artifacts, the students demonstrated their creativity by making up contexts that associate with specific idioms. We analyzed all the 481 photo/sentence sets posted on the wiki and classified them into 12 categories with respect to two dimensions, namely, "types of physical setting" and "types of meaning making". "Types of physical setting" refers to the sources of the physical setting captured by each photo ("natural setting", "physical object manipulation", "human enacted scenario", or "previously published materials" such as book illustrations or TV screenshots). "Types of meaning making" refers to how the associated sentence reflects the student's personal meaning making on the photo (i.e., the relationships between the photo content and the sentence content), which could be "literal meaning making", "extended meaning making" (deductive interpretation) and "creative meaning making" (twisted, creative interpretation). Table 1 features examples of different types of photo/sentence set with the idioms underlined in the original Chinese sentences. To benefit the readers, we translated the Chinese sentences into English with the translations of the idioms underlined.

Here is how we distinguish the personal meaning making types: (1) *Literal meaning making*: The sentence demonstrates a direct interpretation of the photo context – all the elements stated in the sentence are visible in the photos. (2) *Extended meaning making*: The sentence demonstrates a logically deductive interpretation on the photo context – there are elements in the sentence which are invisible in the photos but they are logical deductions from the photo context. For (E), (F), (G), (H) the additional elements are sports games, scoring full marks, a theft, and catching a fish, respectively. (3) *Creative meaning making*: The sentence demonstrates a twisted, perhaps creative re-interpretation on the photo context (i.e., other photo viewers may not interpret the photo in the same way). For example, in photo (I), there is no sign of feeding and eating in the photo and yet the student made up the "plot" of feeding the geese. Sentence (J) turns the photo of two mouse devices into a metaphor as the student imagined that they were animals which collided with each other.

Table 1: Various types of photo contexts and students' meaning making.

| | Natural setting | Physical object manipulation | Human-enacted scenario | Previously published materials |
|---|---|---|---|---|
| Literal meaning making |  (A)这里有两只一模一样的螃蟹。 These two crabs are as alike as two peas. |  (B)那些车横冲直撞，真是不知道他们怎么驾车! Those cars are romping about. I am doubtful about the driving skills of the drivers. |  (C)小青正在睡觉，小蓝不想吵到小青所以她轻手轻脚地走了过去。 Xiaoqing is sleeping. Xiaolan passes her by quietly so as not to disturb her. |  (D)老板怒气冲冲地样子好可怕。 The boss looks fierce when he is in a rage. |
| Extended meaning making |  (E)哥哥平时不爱运动可是却可以在小学的运动会上一鸣惊人地得了第一名 。 My brother does not like sports activities. Yet he excels unexpectedly and came in first the Primary School Sports Meet. |  (F)小明手舞足蹈因为他得了一百分。 Knowing that he scored full marks, Xiaoming dances with joy. |  (G)小明偷了伟德的钱包，伟德火冒三丈对小明拳打脚踢起来。 Knowing that Xiaoming has stolen his wallet, Weide beat Xiaoming up in his rage. |  (H)明明抓到鱼儿时就眉开眼笑。 Mingming grinned from ear to ear when he managed to catch the fish. |
| Creative meaning making |  (I)鹅们津津有味吃我们给的面包。 The geese enjoyed eating the bread we provided with relish. |  (J)它们俩横冲直撞，最后意外发生了! They were romping against each other and in the end resulted in an accident. |  (K)看到工人这样做，非常危险。所以，我一言不发，怕惊吓到她。 Seeing the worker doing a risky act, I tried to be quiet so as not to frighten her. |  (L)整个小岛绿茵环抱，鸟语花香，吸引了很多游人前来度假。 Surrounded by green plants and joyous sceneries, the tiny island has been attracting plenty of tourists to spend holidays there. *(photo source: http://www.pconline.com.cn/)* |

The variety of artifacts reflected the students' greater attention to their surroundings and their more conscious attempts to associate their daily life with the idioms – be it in the school, at home, during family outings, when they read books or watched TV. Their sense of the lack of "natural" contexts for them to take idiom-related photos was compensated by their creativity in manipulating physical objects, enacting situations, or appropriation of relevant published materials. In addition, the post-questionnaire results indicate that the students may have extended, perhaps sub-consciously, their mental "habit" of Chinese-idiom-and-real-life-context association beyond the pilot study, as 75.0% of the students "agree" or "strongly agree" that "after participating in the Chinese idiom learning activities, I think of Chinese idioms more often in my daily life."

### Peer Learning Activities

Besides producing their artifacts, the students carried out two types of peer learning - "learning from peers" and "learning with peers". In terms of "learning from peers", we argue that when a student visits or posts a photo/sentence set on one particular wiki-based "idiom page", she is likely to go through the photos/sentences pertaining to the same idiom made by her peers, which might lead to a mental comparison of the contexts and the grammar among the sentences. That is, the student would either learn from better sentences, or identify and correct her peers' grammatical errors or wrong use of idioms. This may lead to "learning with peers" activities where students comment on and discuss about their peers' contributions. As the statistics presented above indicate the students' more consistent levels of participation in peer learning activities (sentence revisions and comments) as compared with artifact contributions, students who had been less active in sharing photos and sentences may have still engaged in peer learning activities (in the forms of "negotiation of forms" [grammar] and "negotiation of meanings" [idiom usage]) and therefore achieved learning objectives to different extents.

Nevertheless, the students' asynchronous online discussions on the idiom usage were relatively trivial in general. We believed that the lack of training in this aspect was the main reason, as we merely presented a few mock-up photo/sentence artifacts and discussions generated by ourselves to the class at the beginning of the study. As such, we (the researchers and the teacher) applied some online forum facilitation strategies (e.g., Wong & Looi, 2010) by tactfully commenting on student artifacts at the right time and in right ways in order to give space for the students to engage in meaningful discussions. We also advised the teacher to promote class-wide discussions on selected student artifacts during "idiom class" 3 and 4, and eventually conduct in-class consolidation (Activity 4) in the 7$^{th}$ week. The students performed much better when they carried out small-group discussions in Activity 4 as they managed to identify most of the erroneous uses of the idioms on the "idiom pages" assigned to the respective groups and offered good recommendations to improve the sentences. The finding seems to imply that small-group face-to-face discussions would yield better outcomes as compared with asynchronous online discussions and class-wide discussions if the students have not yet acquired the skills of peer reviews – with their relatively low linguistic ability and efficacy being two other possible factors.

We take the idiom page of "东倒西歪" as an example. Table 2 features selected student artifacts posted on the page.

Table 2: Four student artifacts posted on the idiom page of "东倒西歪".

| | | | |
|---|---|---|---|
| <br>(L)这个橱柜上的东西摆得东倒西歪。<br>The objects in the cupboard are rickety. | <br>(M)我的水壶东倒西歪，翻倒了<br>My bottle is rickety and lied down. | <br>(N)哥哥的书柜的书东倒西歪，非常乱！<br>The books in my brother's bookcase are rickety and disorder! | <br>(O)哥哥把书桌弄得东倒西歪。<br>My brother makes his study desk rickety. |

Kovecses & Szabco (1996) define idioms as linguistic expressions whose overall meaning (known as figurative/idiomatic meaning) cannot be predicted from the meanings of the constituent parts (known as literal meaning). For some of the Chinese idioms, however, their applicable contexts have to take the literal meanings into account. "东倒西歪" is one of them. The idiom figuratively means "rickety" or "shaky", and literally means "leaning eastward (right-hand side) and tilting westward (left-hand side)" (usually referring to multiple objects leaning toward multiple directions). When the students discussed about the four photo/sentence sets in Table 2, they agreed that (L) was a correct use of the idiom (there are objects leaning toward both left- and right-hand sides in the cupboard) while (O) was wrong (the paper was messily left on the table, but not "rickety"). Some students proposed a correct alternative idiom "乱七八糟" (messy; at sixes at sevens). On the other hand, both (M) and (N) had generated some debates. In (M), a student argued that it was a wrong use of the idiom as the bottle has already lied down on the table and therefore it is no longer "rickety" or "shaky". However, we re-examined the associated sentence and deduced that it could mean "the bottle had been rickety and now finally lie down", and the photo merely shows the consequence, not the previous state of ricketiness. We found this artifact acceptable, though it is necessary to explain to the students that the idiom does not apply to what the photo itself depicts. Finally, we found the books in photo (N) either stand still or lean toward only one direction (not "eastward and westward"), thus questioning the suitability of applying the idiom. All in all, this is a visual-oriented idiom. The students' photos have helped us in visualizing their understanding (or detecting their misconceptions) in the meaning and applicable contexts of the idiom – something that the conventional ways of assessing them by, for example, sentence making, may not thoroughly achieve. Indeed, we noticed that some of the student artifacts offer seemingly error-free sentences, if they are read independently, but their corresponding photos suggest otherwise (misconceptions in the applicable contexts).

The students' meaningful face-to-face discussions showed that they were on the right track to carry out social meaning making – to bring forward their "individually made" meanings on the idioms in-class and during photo-taking/sentence-making activities to their learning community. Compared with conventional idiom instruction where teachers provide general explanations and demonstrate idiom usage in limited contexts, this personal-to-social meaning making process has effectively pushed the boundary of the learning materials as the students' self-identified contextual uses of the idioms (many of which have caught us by surprise as we have never thought of using the idioms in such contexts) have become a rich resource for the students to perform (essentially inductive) constructivist learning on the idioms.

## Discussion

Researchers have been investigating the facilitation of improvisation and/or creative learner output of various forms as a potentially effective means of language learning (e.g., Hodson, 2008). The mobile technology may support *in situ* improvisation and creative output such as taking appropriate pictures in contexts to illustrate the idioms under study. In our study, while such language learning activities could be carried out without technological support, it is the mobile affordance of *in situ* data collection (in particular, the camera function) that offers them the ease of generating their artifacts and helps others to visualize their idiom-and-context associations. As sharing and "rising above" the shared artifacts are the key to achieve students' deep learning of the idioms, the incorporation of the Web 2.0 (wiki) technology further enhances their social learning space by "affording" them rapid artifact revisions and interactions.

This learning process design is grounded in several existing learning language approaches and yet with some novelty. It is inductive vocabulary learning, yet not entirely based on teacher-supplied resources (e.g., Mishan, 2004), but student-generated ones, which are more authentic to the students' daily life. It is language learning by meaning making, yet not through reading or conversational comprehension (e.g., Donato, 1994) but students' idiom-context associations. It emphasizes contextual learning, productive outputs and socio-constructivism. In addition, it reinforces a principle of learning – make errors work for the students and not against them (Rubin & Thompson, 1982). In our study, an "idiom page" that contains both the correct and ambiguous (or erroneous) idiom usages would turn out to be an excellent venue for student discussion. The teacher should hold back his corrective comments to encourage debates among students on diversified views.

The pilot study has also informed us that pertaining to personalized learning, the student artifacts show some indicators of personalized seamless learning, i.e., students' proactive association of what they have learnt in-class with what they are experiencing in daily life, and perhaps with further enactment, reflection and personal meaning making. Pertaining to peer learning, the rich results of the social meaning making practice had facilitated ongoing student discussions and inductive comparisons of varied student-identified contexts. Together with the activities of watching idiomatic animations, this seamless language learning process fulfils Nation's (2001) three-stage process for (essentially personalized) vocabulary learning, and even extends it to involve social learning, and turns the language subject into an authentic learning experience.

On the other hand, the challenges faced in our pilot study have helped us in identifying the implementation. Our original intention of leveraging on general digital natives' eagerness to share their real-life experiences 'on-the-fly', mediated by Web 2.0 (Prensky, 2004) and mobile devices, did not take off. The students were keener on game-playing than blogging. The findings will trigger us to apply additional strategies to prepare the students who will involve in the full-scale study in the following year.

## Conclusion

In this paper, we report a pilot study within the MALL paradigm that is both creative learner output-centric and seamless learning-inclined. These two areas, which promised great educational potential, have been seriously untapped in prior MALL studies. In our study, from artifact generation to peer learning activities, we observed a trajectory of noticing-retrieving-generating process, and personal-to-social meaning making among the students, which we wish to analyze further in order to unveil its relation with Chomsky's (1986) distinctions between Internal- and External-language learning, and the general thinking skills involved (such as that with respect to Bloom's Taxonomy). While the results of the study are promising, much work need to be done to ensure the students' learning motivation and enthusiasm at the in-class mobile assisted learning could be extended to their off-campus time when they are assigned handhelds 24x7. Still, with proper design and implementation of seamless language learning, we envisage the potential of MALL to revolutionize language learning by using mobile devices as individual students' personal "learning hubs", and further synergize the formal (in-class) and informal (real-life), as well as the personal and social language learning spaces.

## References


Bolte, H., & Herrlitz, W. (1986). Language learning processes and strategies in the foreign language classroom: A reconstructive and interventive model. In G. Kasper (Ed.), *Learning, Teaching and Communication in the Foreign Language Classroom* (pp. 195-223), Aarhus: Aarhus University Press.

Bourke, J. M. (2005). The grammar we teach. *Reflections on English Language Teaching, 4*, 85-97.

Chomsky, N. (1986). *Knowledge of Language. It's Nature, Origin and Use*. New York: Praeger.

Deng, C. (2001). The common mistakes in the usage of idioms and the solutions. *Learning Language, 1*, 40-41.

Donato, R. (1994). Collective scaffolding in second language learning. In: J.P. Lantos & G. Appel (Eds.), *Vygotskian Approaches to Second Language Research* (pp.33-56), Norwood, NJ: Ablex.

Elliot, R. T., & Zhang, Q. (1998). Inference in learning context-dependent words. *Educational Psychology, 18*(1), 5-25.

Gass, S. M. (1997). *Input, Interaction, and the Second Language Learner.* Mahwah, N.J.: Lawrence Erlbaum.

Hedley, J. (1992). Surviving to spaek new language: Mary Daly and Adrienne Rich. *Hypatia, 7*(2), 40-62.

Hodson, R. J. (2008). Media humour in the language classroom. *Proceedings of CLaSIC 2008 (pp.669-679)*, Singapore.

Hoopingarner, D. (2009). Online language learning 2.0: new tools, new paradigms. *Proceedings of ICICE 2009* (pp.57-64), Taipei, Taiwan.

Jonassen, D. H. (1991). Objectivism vs constructivism: do we need a new educational paradigm?. *Educational Technology, 9*, 5-14.

Kauffman, S. (1995). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford: Oxford University Press.

Kovecses, Z. & Szabco, P. (1996). Idioms: a view from cognitive semantics. *Applied Linguistics. 17*(3), 326-355.

Kukulska-Hulme, A., & Shield, L. (2007). An overview of mobile assisted language learning: can mobile devices support collaborative practice in speaking and listening?. *EuroCALL 2007*.

Kumaravadivelu, B. (1994). The postmethod condition: (E)merging strategies for second/foreign language learning. *TESOL Quarterly, 28*(1), 27-48.

Levy, M., and Kennedy, C. (2005). Learning Italian via mobile SMS. In: A. Kukulska-Hulme and J. Traxler *(Eds.)*, *Mobile Learning: A Handbook for Educators and Trainers*. London: Taylor and Francis.

Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 255-270), Cambridge: Cambridge University Press.

Locke, J. L., and Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, *29*, 259-325.

Long, M. H. (1985). Input and second language acquisition theory. In: S.M. Gass & C.G. Madden (Eds.), *Input in Second Language Acquisition* (pp.377-393), Rowley, MA: Newbury House.

Looi, C. K., Seow, P., Zhang, B., So, H.-J., Chen, W., & Wong, L. H. (2009). Leveraging mobile technology for sustainable seamless learning: a research agenda. *BJET*. Published online: 6 February, 2009.

Miller, G. A., & Gildea, P. M. (1987). How children learn words. *Scientific American. 257*, 94-99.

Mishan, F. (2004). Authenticating corpora for language learning. *ELT*, *58*(3), 219-227.

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge, UK: Cambridge University Press.

Ogata, H., Akamatsu, R., & Yano, Y. (2004). Computer supported ubiquitous learning environment for vocabulary learning using RFID tags. *Proceedings of TEL 2004* (pp.121-130), Toulouse, France.

Prensky, M. (2004). The emerging online life of the digital natives. Retrieved Sep 21, 2009, from: http://www.marcprensky.com/writing/Prensky-The_Emerging_Online_Life_of_the_Digital_Native-03.pdf

Rubin, J., & Thompson, I. (1982). *How to Be a More Successful Language Learner.* Boston: Heinle and Heinle.

Sagvinon, S. J. (1991). Research on the role of communication in classroom-based foreign language acquisition: On the interpretation, expression, and negotiation of meaning. In: B.F. Freed (Ed.), *Foreign Language Acquisition Research and the Classroom* (pp.31-45). Lexington, MA: D.C. Health.

Salaberry, M. R. (1996). A theoretical foundation for the development of pedagogical tasks in computer mediated communication. *CALICO, 14*(1), 5-36.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *Modern Language, 79*, 15-28.

Stevick, E. (1996). *Memory, Meaning and Method: A View of Language Teaching. (2nd ed.)*. Boston: Heinle and Heinle.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In: S. Gass & C. Madden (Eds.), *Input and Second Language Acquisition* (pp.235-252*).* Rowley, Mass: Newbury House.

Thornbury, S. (2001). *Uncovering Grammar*. Oxford: Macmillan Heinemann.

Thorton, P., and Houser, C. (2002). M-learning in transit. In: P. Lewis *(Ed.)*, *The Changing Face of CALL* (pp. 229-243). Lisse, The Netherlands: Swets and Zeitlinger.

Titone, R. (1969). Guidelines for teaching second language in its own environment. *Modern Language*, *53*(5), 306-309.

Webb, S. (2005). Receptive and productive vocabulary learning: the effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*, 33-52.

Widdowson, H. G. (1978). *Teaching Language as Communication*. Oxford: OUP.

Winters, N. (2007). periLEARN: contextualised mobile learning in the era of Web 2.0. *Proceedings of IADIS Int'l Conference on Mobile Learning 2007* (pp. 89-95), Lisbon, Portugal.

Wong, L. H., and Looi, C. K. (2010). Online discussion and e-mentoring strategies in continuing education courses. In: E.M.W. Ng (Ed.), *Comparative Blended Learning Practices and Environments* (pp. 146-169). Hershey, PA: IGI.

Zhang, B., Wong, L. H., Seow, P., Chen, W., & Looi, C. K. (2009). 'Mobilizing' curriculum materials for seamless learning: opportunities and challenges. *Proceedings of GCCCE 2009* (pp.66-71), Taipei, Taiwan.

# Group Micro-creativity in Online Discussions: Effects of New Ideas and Social Metacognition

Gaowei Chen, The Hong Kong Institute of Education, Hong Kong, China, gaoweichen@gmail.com
Ming Ming Chiu, State University of New York – Buffalo, New York, USA, mingmingchiu@gmail.com
Zhan Wang, The Hong Kong Institute of Education, Hong Kong, China, janjanwang@gmail.com

**Abstract**: This study examines how group members' new ideas and social metacognition in recent messages (micro-time context) affected a current message's micro-creativity (correct, new idea) during online discussions. Dynamic multi-level analysis was used to model statistically 894 messages by 183 participants on 60 high school mathematics topics from one of the world's largest mathematics problem solving website (www.artofproblemsolving.com). Results showed that new ideas (correct, new ideas and justifications) and social metacognition (correct evaluations and questions) in recent messages increased the likelihood of a current message's micro-creativity. Applied to practice, these results suggest that teachers might increase students' micro-creativity by encouraging them to post more correct, new ideas, justify their own ideas, evaluate others' ideas carefully, and ask more questions during online discussions.

## Introduction

Students are increasingly using online asynchronous discussions for learning (Tallent-Runnels, Thomas, Lan, Cooper, Ahern, & Shaw, 2006), in part because online discussions are more accessible and more equal compared to traditional face-to-face (FTF) discussions (Harasim, 1993). Also, online discussions allow more time for students to prepare, do higher order and critical thinking, and search for extra information before posting a message in the discussion (Pena-Shaff & Nicholls, 2004). By understanding how online discussions evolve message by message, educators can improve their quality and facilitate students' online learning.

Past studies of face-to-face (FTF) discussions have shown that recent group processes (micro-time context) can affect the likelihood of a current speaker's creation of correct, new ideas (*micro-creativity*). In this paper, creativity refers to the "small c" creativity of ordinary people in daily life, not the "big C" creativity of new knowledge or products that substantially affect society (Gruber & Wallace, 1999). Hence, micro-creativity is defined as an expressed *correct, new idea* relative to the group members' experiences during a discussion.

Chiu (2008) showed that recent speakers' justifications and correct evaluations of one another's ideas aided a current speaker's micro-creativity. This raises the issue of how recent phases and messages (local message context) affect a current message's micro-creativity during online discussions. As micro-creativity is central to correct solutions and problem solving success, understanding the group processes that affect micro-creativity can help educators improve students' online discussions.

Most previous studies of online discussions examined the *individual* properties of each discussion message from course related forums (e.g., Hara, Bonk, & Angeli, 2000). Their results showed that many students processed course information at high cognitive levels during online discussions, supporting the claim that online discussion can promote students' micro-creativity. However, researchers have not systematically examined the *relationships* among messages to characterize online moment-to-moment creative processes.

In this study, a step is taken in this direction by examining how new ideas, justifications, social metacognition, and individual characteristics in recent messages facilitate or hinder a current message's micro-creativity during online mathematics problem discussions. This study contributes to the research literature in three ways. First, we introduced hypotheses regarding how new ideas, social metacognition, and individual characteristics help create a local message context that might influence micro-creativity during online discussions. Second, we explicated a new statistical discourse analysis method to model multi-topics of online conversations (Chiu & Khoo, 2005). Lastly, we applied this new method to analyze 60 high school-level mathematics topics from a non-course related forum in one of the world's largest mathematics problem solving website (AoPS, www.artofproblemsolving.com). By understanding the group processes that affect micro-creativity, group members can work together more creatively during online discussions.

## Theoretical Perspective

In this study, several hypotheses are introduced regarding how local message context might influence micro-creativity, which in turn might affect problem solving success during online discussions (see Figure 1). In this paper, "E-poster" is defined as the author of an online discussion message.

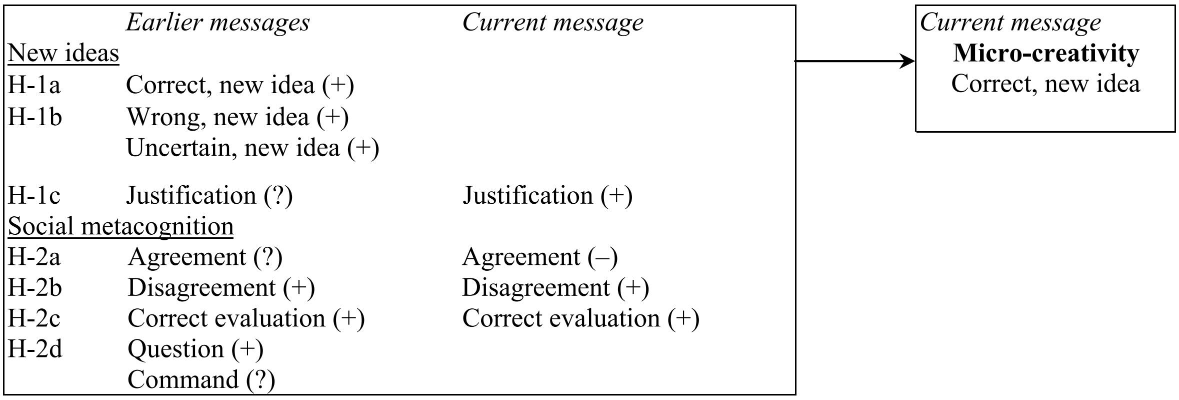



Figure 1. Model of recent messages' effects on the outcome variable micro-creativity in a current message. (Symbols in parentheses indicate the expected direction of the relationship with the outcome variable: positive [+], negative [–], or unknown [?]). H = hypothesis.

## New Ideas

In FTF or online discussions, group members often have diverse perspectives and sources of knowledge (Swann, Kwan, Polzer, & Milton, 2003). Capitalizing on this diversity, heterogeneous group members often contribute new ideas that have not been mentioned earlier in the discussion. Like FTF discussants, e-posters' new ideas can be correct, wrong, or uncertain in many contexts (e.g., high school mathematics). Suppose a group is working on a mathematical problem (e.g., "What is 7 times 9?"). An e-poster's new idea can be correct if it does not violate any mathematics or problem constraints ("7 times 9 is 63"), wrong if it violates at least one mathematics or problem constraint ("7 times 9 is 64"), or uncertain if its validity cannot be determined by considering mathematics or problem constraints ("I'm not sure, but this link might help, [hyperlink]").

Ideally, e-posters build micro-creativity directly on previous micro-creativity. For example, in responding to a micro-creativity message, "7×9 = 7×[10-1]", an e-poster might continue the thread and build new micro-creativity: "7×[10-1] = 7×10-7×1 = 70-7 = 63." Or, e-posters might not be able to build micro-creativity directly based on previous micro-creativity. Instead, they might test the previous micro-creativity against their established facts, existing cognitive schema, or personal experience, especially in online discussions fostering critical thinking (Gunawardena, Lowe, & Anderson, 1997). By testing the micro-creativity, group members are likely to identify the obstacle(s) in it. Then, they can either use questions (e.g., "Could you explain the steps further?") or disagreements (e.g., "I don't think your steps are complete") to express the obstacle and invite further elaborations or justifications (Perturbations; Piaget, 1985). Then, in responding to the questions or disagreements, group members are likely address the obstacle and make micro-creativity (e.g., "Here are the full steps, 7×[10-1] = 7×10-7×1 = 70-7 = 63"). Hence, micro-creativity might spark micro-creativity in subsequent messages.

E-posters might also post wrong, new ideas during online discussions. A group of e-posters' diversity and increased critical thinking might help them detect flaws, modify these wrong ideas, and make micro-creativity (De Lisi & Goldbeck, 1999). However, wrong, new ideas might also lead a group astray, so the effect of them on micro-creativity is unclear.

Lastly, as online discussants are often from different cities or countries, they are likely to have diverse views and sources of knowledge, thereby increasing the likelihood of presenting uncertain, new ideas, which can be uncertain resources ("I'm not sure, but this link might help, [hyperlink]"), tentative suggestions ("There might be a general way to solve such problems"), and personal preference ("I think this method is easier to understand"). The useful part in these uncertain, new ideas might help elicit micro-creativity. However, like wrong, new ideas, these uncertain, new ideas might also lead a group astray. So the effect of them on micro-creativity is unclear as well.

E-posters' ideas might or might not be justified. A justified idea demonstrates that the proposer has tried to justify his or her claim (e.g., "7×9 is 63") by linking it to data/warrants (e.g., showing reasoning/evidence, "because 7×9 =7×[10-1] = 7×10-7×1 = 70-7 = 63"), backings (e.g., "we can solve it by converting multiplication to addition"), or a combination of them (Guiller, Durndell, & Ross, 2008; Toulmin, 2003). A justification argues that an idea is reasonable and can often improve the validity of it (Lindow, Wilkinson, & Peterson, 1985). Furthermore, justifying an idea in written form might help identify errors and provide clearer and more precise steps. Hence, justifications support the validity of an idea and can increase micro-creativity.

H-1a. Earlier correct, new ideas facilitate micro-creativity.
H-1b. Earlier wrong, new ideas and uncertain, new ideas facilitate micro-creativity.

H-1c. Justification in a message facilitates its micro-creativity.

### Social Metacognition
Successful group problem solving often involves social metacognition in the problem content space (Roschelle, 1992). Whereas individual metacognition is monitoring and controlling one's own knowledge, emotions, and actions, social metacognition is group members' monitoring and control of one another's knowledge, emotions, and actions (Chiu & Kuo, 2009). Group members use social metacognitive strategies to evaluate one another's ideas (agreement, disagreement, or neutral), recognize problems, and invite audience participations (question, command, or statement).

In responding to a previous message, a responder might have supportive information, conflicting information, or both supportive and conflicting information. With supportive information, the responder is likely to agree with the proposer's idea and post a supportive message (e.g., "Yes, you are right, I got the same answer"). In such cases, a new idea, which might conflict with the old one, is less likely to be voiced. Hence, an agreement is less likely to aid micro-creativity.

Or, a responder might have conflicting information, which often indicates an inconsistency in understanding of terms, concepts, or schemas (Carson, Butcher, & Coleman, 1988). Then, the responder is likely to disagree with the proposer's idea and respond with a conflicting message (e.g., "No, you are wrong, 7×9 is not 64") to identify the inconsistency. According to *socio-cognitive conflict theory* (Piaget, 1985), the disagreements might aid micro-creativity, either immediately or in subsequent messages.

Lastly, an e-poster might have both supportive and conflicting information toward a previous idea. In online discussion reduces face concerns and allows delayed responses, which likely free an e-poster to provide the conflicting information as well. In such cases, the responder is more likely to disagree with the proposer's idea on at least one point (e.g., "Yes, your answer is correct, *but the method is too complex*"). Again, according to socio-cognitive conflict theory, the disagreement is likely to aid group members' micro-creativity.

Group members' evaluations of one another's ideas can be correct or wrong. Correct evaluations support previous correct ideas by providing supportive information ("Yes, 63 is right") or identify flaws in previous wrong ideas by providing conflicting information ("No, 7×9 is not 64"). In contrast, biased evaluations reject correct ideas ("Nope, 7×9 is not 63") or accept flawed ones ("I agree, 7×9 is 64"), embedding flaws in their partially shared understandings. Group members using these wrong, shared understandings might carry these flaws into their new ideas, resulting in more wrong ideas and fewer correct ones (Chiu, 2008).

H-2a. Agreement in a message hinders its micro-creativity.
H-2b. Disagreements in the current and earlier messages facilitate micro-creativity.
H-2c. Correct evaluations in the current and earlier messages facilitate micro-creativity.

Online discussions can promote e-posters' thoughtful questions and micro-creative answers through its asynchronous discussion mode and increased critical thinking (Gunawardena et al., 1997). The question-answer interactions, when enhanced by critical thinking, are more likely to generate thoughtful questions and micro-creativity. Unlike questions, commands stop or demand stop or demand audience participation and typically begin with verbs. Compared to FTF commands, online commands' effect on group members' micro-creativity is likely smaller. The anonymous nature of online discussions helps reduce a command's face threat. As a result, the target e-poster is less likely to retaliate emotionally and more likely to respond rationally to the command (e.g., "Explain your steps!") and create a new idea (e.g., "Here are my steps…"). Thus, commands might or might not reduce micro-creativity in online discussions.

H-2d. Earlier questions facilitate micro-creativity.

Other factors such as individual characteristics (online displayed gender, past posts, initiator of topic), message # (greater # indicates later posting) and message length (the number of words per message) were entered as control variables.

## Method

### Participants
In this study, 183 e-posters discussed 60 mathematics problems in the High School Basics (HSB) forum. About 80% e-posters were 13- to 18-year-old high school students in grades 9-11 from all around the US (calculated approximately from the participants' displayed personal information). The other 20% e-posters were from countries worldwide. The HSB forum is a free, non-course-related forum from the *Art of Problem Solving* website, which is among the world's largest mathematics problem solving websites.

## Data

Sixty problems were randomly selected from the HSB forum of AoPS, excluding problems with less than 4 reply messages. The 60 problems covered the subjects of algebra, geometry, number theory, and counting. Every problem had definite, correct solution(s). The number of responses to the 60 problems ranged from 6 to 26 and the discussion durations of them ranged from 4 to 285 hours.

Each problem and its reply messages were linked to one another by multiple threads, single connections, and quotes of previous messages. See Figure 2 for example relationships between a problem and its 12 responses. The left side outlines the positions of the problem and its 12 responses on the web page. The right side shows their relationships. The number "0" denotes the initial problem; "1" through "12" indicates 12 reply messages in the order of time, where "1" means the earliest reply, "12" the latest.

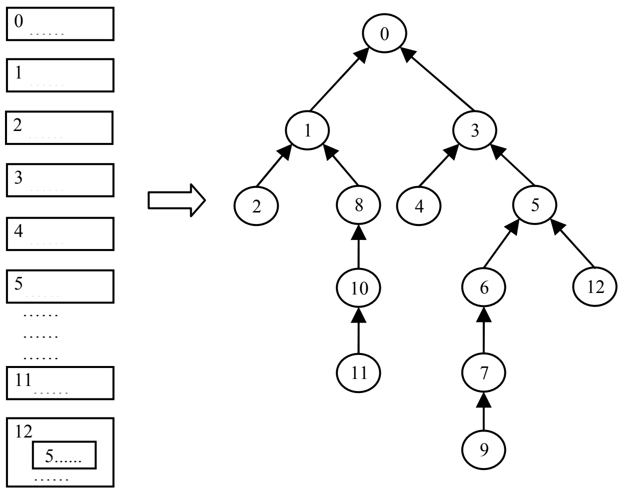



Figure 2. The relationships between a problem and its reply messages.

*Multiple threads*. The problem and 12 reply messages occurred along five discussion threads: (a) 0→ 1 → 2, (b) 0 → 1 → 8 → 10 → 11, (c) 0 → 3 → 4, (d) 0 → 3 → 5 → 6 → 7 → 9, and (e) 0 → 3 → 5 → 12. Messages in each thread were ordered by time, but they were not necessarily consecutive. In thread (d) for example, message #9 followed message #7 (not #8) and message #5 followed message #3 (not #4).

*Single connections*. All reply messages were linked together by single connections. The forum's interface design constrained each message to respond to only one previous message, which helped form clear multiple discussion threads and avoid ambiguous relationships among messages. For example, message #9 only responded to message #7.

*Quote of previous messages*. For two messages that were not consecutive along each single connected thread, the reply message automatically quoted the message to which it responded. By doing so, both messages can be read at the same place. As illustrated in Figure 2, message #12 quoted message #5, with the latter enclosed in the indented rectangle. For two consecutive messages that can be read at same computer window, the reply message did not necessarily quote the previous one.

## Variables

Variables for a single message included the e-poster's past posts, message #, message length, and the following binary variables: micro-creativity (correct, new idea), wrong & new idea, uncertain & idea, justification, repetition, agreement, disagreement, correct evaluation, question, command, masculine, feminine, and initiator of the discussion topic. The analyses use two sets of variables: *current variables* measuring properties of the current message (0) and *lag variables* measuring properties of earlier messages in the same thread (-n, where n = 1, 2, 3, 4, …).

## Coding

Two students coded each message separately. Then, they settled all coding disagreements by consensus. All 894 messages in the transcript were included and coded. Krippendorff'sα (2004) was used to compute the inter-rater reliability. Unlike other inter-rater reliability measures, Krippendorff'sα applies to any number of coders, any number of categories or scale values, any level of measurement, any sample size, and incomplete data. Its values range from -1 (maximum disagreement) to 1 (perfect agreement). A value near 0 indicates chance agreement among the coders, and a value of 0.7 or higher indicates satisfactory agreement.

The web server automatically recorded each e-poster's number of past posts and each message's # and length. Other variables were coded as binary variables. A message's variable (-n) was coded as the values of the variable of the message n connections prior. For example, in Figure 2, the value of wrong & new idea (-1) for message #9 was the same value of wrong & new idea (0) for message #7, since #7 was #9's previous message.

## Analyses

*Dynamic multi-level models* (Chiu & Khoo, 2005) separate unexplained error into message (level one) and topic (level two) components, thereby removing the correlation among error terms resulting from messages nested within topics. For the outcome variable micro-creativity, we first added *s* topic level variables as control variables: algebra topic, geometry topic, and number theory or counting topic (**S**). We then added *t* current message variables (0) as control variables: masculine, feminine, past posts, initiator, message # and message length (**T**). As the likelihood ratio test for significance of additional explanatory variables was not reliable for this estimation method, Wald tests were used. Non-significant variables and interactions were removed.

We then added *u* current message variables (0): justification, agreement, disagreement, and correct evaluation (**U**). Likewise, we repeated the procedure for **T** on **U**. Then, we tested for interaction effects among pairs of significant variables in **U**. Non-significant variables and interactions were removed from the specification. Next, we tested if the *u* regression coefficients (($\beta_{uj} = \beta_{u0} + f_{uj}$) differed significantly at the topic level ($f_{uj} \neq 0$?). If yes, we kept these additional parameters in the model. Otherwise, we removed them.

Next, we tested the hypotheses by entering lag variables measuring the property of earlier messages (-n), first lag variables (-1), then lag variables (-2), lag variables (-3), and lastly, lag variables (-4). First, we added lag variables (-1) at the message level: message length (-1), masculine (-1), feminine (-1), past posts (-1), initiator (-1), correct, new idea (-1), wrong & new idea (-1), uncertain & new idea (-1), justification (-1), repetition (-1), agreement (-1), disagreement (-1), correct evaluation (-1), question (-1), and command (-1) (**V**). Likewise, we applied the procedure for **U** on **V**. Then, we repeated the procedure for lags -2, -3, and -4 of the variables in **V**.

We used an alpha level of .05 for all statistical tests. Benjamini, Krieger, and Yekutieli's (2006) two-stage linear step-up procedure was used to control for the false discovery rate. We used Higgins and Thompson's (2002) $I^2$ index, which assesses both the statistical significance and the extent of heterogeneity simultaneously, for testing serial correlation in the residuals of the regressions for all topics.

Based on the multi-level analysis results, the path analysis estimated the direct and indirect effects of the significant explanatory variables separately to compute their total effects (Kennedy, 2003). To facilitate the interpretation of these results, we converted the total effects of each predictor to odds ratios, indicated by the percentage increase or decrease (+ X% or – X%) in the likelihood of an outcome variable (Judge, Griffiths, Hill, Lutkepohl, & Lee, 1985).

## Results

Coding of each dimension showed high inter-rater reliability. The inter-rater reliabilities as measured by Krippendorff's α for knowledge content, evaluation, and invitational form were 0.85, 0.89, and 0.92 respectively (corresponding percentages of agreement were 90%, 94%, and 97%). The multi-level variance components analysis showed that e-posters' micro-creativity did not differ significantly across topics, so single-level analyses at the message level were adequate. No interaction effects were detected in the analyses.

As variables measuring earlier messages required data from preceding turns, some messages could not be used, leaving fewer messages available for analyses. As a variable measuring messages up to two turn prior was significant (i.e., correct, new idea [-2]), 501 messages by 137 e-posters were included in the final analyses.

Correct, new idea in the previous message (-1) reduced the likelihood of micro-creativity (–5%: 9% → 4%; when a previous correct, new idea did not occur, micro-creativity occurred 9% of the time; when a previous correct, new idea occurred, micro-creativity occurred 4% of the time; see Table 1 and Figure 3), whereas correct, new idea two messages prior (-2) increased it (+11%: 9% → 20%), supporting H-1a. Justifications aided micro-creativity (+48%: 9% → 57%), showing that a justified, new idea was more likely to be correct, supporting H-1c.

Table 1: Total effects of each explanatory variable on micro-creativity and explanatory variables.

| Explanatory variable (E) | | Target (T) | P(T│E) (%)[a] | P(T│~ E) (%)[b] | Effect (%) |
|---|---|---|---|---|---|
| (1) | Past posts (0) | Micro-creativity | 11 | 9 | +2[c] |
| (2) | Initiator (0) | Micro-creativity | 4 | 9 | –5 |
| (3) | Justification (0) | Micro-creativity | 57 | 9 | +48 |
| (4) | Correct evaluation (0) | Micro-creativity | 30 | 9 | +21 |
| (5) | Question (-1) | Micro-creativity | 21 | 9 | +12 |
| (6) | Correct, new idea (-1) | Micro-creativity | 4 | 9 | –5 |
| (7) | Correct, new idea (-2) | Micro-creativity | 20 | 9 | +11 |
| (8) | Question (-1) | Justification (0) | 43 | 22 | +21 |

| | | | | | |
|---|---|---|---|---|---|
| (9) | Question (-1) | Correct evaluation (0) | 15 | 36 | –21 |
| (10) | Correct, new idea (-2) | Question (-1) | 55 | 25 | +30 |
| (11) | Correct, new idea (-2) | Correct, new idea (-1) | 19 | 39 | –20 |

[a] Probability that the target occurs, given that the explanatory variable does occur.
[b] Probability that the target occurs, given that the explanatory variable does *not* occur.
[c] The +2% effect is explained by a 100% increase above the mean of e-posters' past posts.

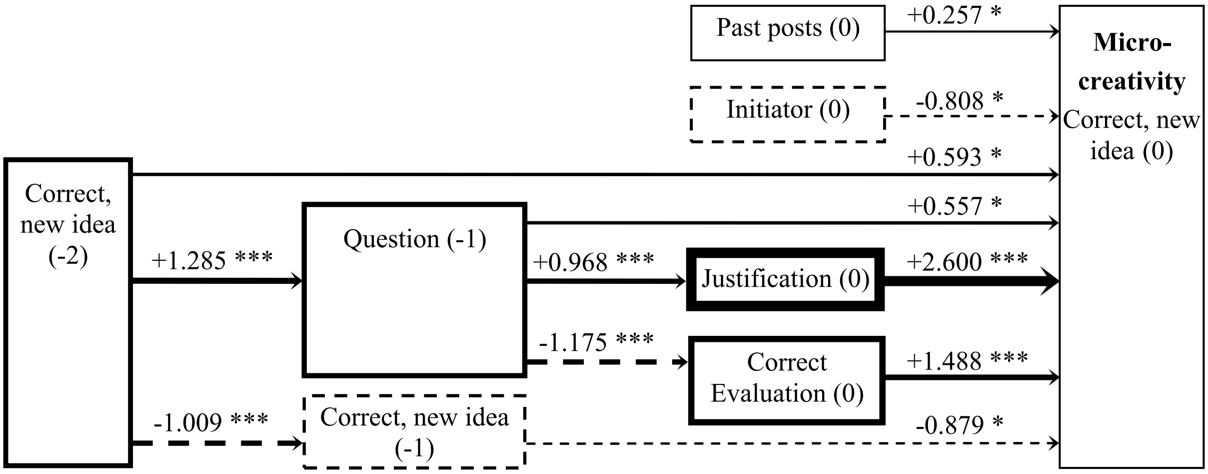



Figure 3. Path analysis of significant explanatory variables of micro-creativity using two-level Logit. Values are standardized parameter coefficients. Explanatory variables inside solid boxes indicate positive overall effects, while those inside dashed boxes indicate negative overall effects. Solid arrows (→) indicate positive direct effects, while dashed arrows (---→) indicate negative direct effects. Line widths indicate effect sizes according to the percentage changes in Table 1.

Correct evaluation (0), i.e., a correct evaluation in the current message, increased micro-creativity (+21%: 9% → 30%), supporting the first half of H-2c. If evaluating a previous idea correctly, either agreeing with a correct idea or disagreeing with a wrong idea, an e-poster was 21% more likely to show micro-creativity.

A question in the previous message (-1) increased micro-creativity (+12%: 9% → 21%), supporting H-2d. Commands in the earlier messages did not affect micro-creativity. As the data had only two commands (out of 501 messages), the statistical power was too low to test their effects.

This Logit model had an 82% accuracy rate for predicting micro-creativity in any given message ($y_{ij}$* vs. $y_{ij}$). The corresponding Probit model produced consistent parameter estimates. Furthermore, the final model' $Q$ statistics and $I^2$ index showed no significant serial correlation of residuals for the 60 topics up to lag 3 ($df$ = 60, 116, 170, respectively; $Q$ = 46.062, 89.786, 106.139, respectively; p>.05 for all; $I^2$ index = 0 for all). So, the time-series model was likely appropriate.

## Discussion

Of all 501 messages, 64% had new ideas, including 31% micro-creativity, 7% wrong, new ideas, and 26% uncertain, new ideas. The high proportion of new ideas is consistent with Heckman and Annabi's (2005) study where 83% online discussion messages and 30% FTF utterances contained new ideas (e.g., analysis, integration). Also, micro-creativity occurred more often than wrong, new ideas or uncertain, new ideas in this study, consistent with Chiu's (2008) study of FTF discussions showing that correct ideas occurred 20% of the time while wrong ideas occurred 10% of the time.

### Correct, New Ideas Increases Subsequent Micro-creativity

Correct, new ideas in earlier messages aided a current message's micro-creativity. Specifically, Correct, new ideas in the previous message reduced the likelihood of a current message's micro-creativity, whereas correct, new ideas two messages prior increased it. Overall, correct, new ideas in earlier messages increased the likelihood of a current message's micro-creativity.

These results suggest that, in responding to correct, new ideas, e-posters were less likely to build micro-creativity directly. Instead, they tended to ask questions (or perhaps use other perturbations) as the path analysis showed that correct, new ideas predicted questions in this study. Then, in responding to these questions, they were likely to address the obstacle(s) in the earlier correct, new ideas and make new micro-creativity. Hence, correct, new ideas helped elicit more micro-creativity in subsequent messages, showing that the chain reactions of micro-creativity might occur in online discussions, thereby supporting the claim that online discussions can promote students' micro-creativity.

The effects of wrong, new ideas and uncertain, new ideas on micro-creativity remained unclear. The useful parts in wrong or uncertain messages might have aided e-posters' micro-creativity (Nijstad, Diehl, & Stroebe, 2003). However, the wrong or uncertain information in these messages might also lead e-posters astray, thereby lowering the efficiency in eliciting micro-creativity.

Together, these results suggest that micro-creativity were more important than wrong or uncertain, new ideas for eliciting micro-creativity in this study, unlike Chiu's (2008) study of FTF discussion showing that wrong, new ideas were more important than micro-creativity for eliciting micro-creativity.

## Justifications Increase Immediate Micro-creativity

Justifications aided immediate micro-creativity, consistent with the studies of FTF discussions (e.g., Chiu, 2008; Goldbeck, 1998; Lindow et al., 1985). Furthermore, justifications had the strongest effect on micro-creativity (+48%), larger than the combined effects of all other predictors. This result highlights the importance of justifications in aiding micro-creativity during online discussions. The process of justifying a new idea might help identify flaws in each step, obtain precise reasoning steps, and lead to micro-creativity. Hence, e-posters can capitalize on the time flexibility of online discussions to think and search for evidence to justify their new ideas, thereby aiding micro-creativity.

## Correct evaluations Increase Immediate Micro-creativity

Like justifications, correct evaluations also aided immediate micro-creativity. If evaluating a previous idea correctly, the e-poster was likely to show micro-creativity. Controlling for correct evaluations, agreements or disagreements alone did not affect e-posters' micro-creativity. These results show that the accuracy of e-posters' evaluations was more important than the type of evaluations for facilitating micro-creativity in this study. Recognizing the previous ideas as correct or wrong helps e-posters build on them to make micro-creativity (Barron, 2003). Hence, e-posters can take time to evaluate one another's ideas carefully, thereby increasing the likelihood of presenting correct evaluations that aid immediate micro-creativity.

During the online discussions in this study, agreements did not reduce the likelihood of micro-creativity, nor did disagreements increase it, possibly because e-posters used fewer simple agreements and more simple disagreements. Unlike FTF discussions, online discussions do not expect any participant to respond soon or respond at all. Hence, simple confirmations or false agreements without micro-creativity (e.g., the "Uh-huh" or "Yeah" in FTF discussions) are less likely to occur in online discussions. Instead, the reduced face concern and increased psychological distance of e-posters might incline them to disagree more often with others' ideas, even without sufficient evidence for the opposition (e.g., "You're wrong", "I disagree"; Reinig & Mejias, 2004). Hence, simple disagreements without micro-creativity are more likely in online discussions. As a result, the expected negative effect of agreements and positive effect of disagreements on micro-creativity are likely to be mitigated.

## Questions Increase Subsequent Micro-creativity

Questions yielded greater micro-creativity, showing that questions often received satisfactory responses. Questions also increased the likelihood of micro-creativity through inviting justifications (see Figure 3). These results show that e-posters were likely to ask questions that invite explanations or justifications, supporting the view that online discussions can promote e-posters' higher order thinking (Schrire, 2004).

Questions often occurred after micro-creativity and helped elicit more micro-creativity. After micro-creativity, the responders often asked questions inviting further explanations or justifications, which in turn yielded micro-creativity. The "micro-creativity (-2) → question (-1) → micro-creativity" and "micro-creativity (-2) → question (-1) → justification (0) → micro-creativity" effects suggest that e-posters are likely to identify the obstacles in micro-creativity, express them by questions, and use micro-creativity to address the questions. Meanwhile, these effects highlighted the importance of questions in facilitating micro-creativity during online discussions.

## Implications for Teachers and Students

This study highlights the importance of micro-creativity, justifications, correct evaluations, and questions in facilitating micro-creativity during online discussions. To aid students' online micro-creativity and problem solving, teachers can encourage them to post more micro-creativity, justify their own ideas, evaluate one another's ideas carefully, and ask more questions during online discussions.

First, teachers can encourage students to post more micro-creativity during online discussions, which is not necessarily the ultimate solution to the problem. As micro-creativity fostered subsequent micro-creativity, even minor micro-creativity might serve as bases for creating more comprehensive micro-creativity. Second, teachers can encourage students to justify their own ideas and evaluate others' ideas carefully. Supported by a justification, a new idea is more likely to be reasonable and micro-creative. Meanwhile, by evaluating a previous idea carefully, a student is more likely to evaluate it accurately and correctly, and thereby build micro-creativity on it. Third, teachers can encourage students to ask more questions during online discussions. By asking questions that invite explanations or justifications, students can stimulate one another's higher order thinking and micro-creativity. Lastly, teachers can intervene to ask leading questions to inspire students' micro-creativity.

## References


Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*, 491–507.

Carson, R., Butcher, J., & Coleman, J. (1988). *Abnormal psychology in modern life* (8th ed.). Glenview, IL: Scott, Foresman.

Chiu, M. M. (2008). Flowing toward correct contributions during group problem solving: A statistical discourse analysis. *Journal of the Learning Sciences, 17*(3), 415–463.

Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multi-level analysis. *Small Group Research, 36*, 1–32.

Chiu, M. M., & Kuo, S. W. (2009). Social metacognition in groups: Benefits, difficulties, learning, and teaching. In C. B. Larsen (Ed.), *Metacognition: New research developments*. New York: Nova Science Publishers.

De Lisi, R., & Goldbeck, S. L. (1999). Implications of Piagetian theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 3–37). Mahwah, NJ: Erlbaum.

Goldbeck, S. (1998). Peer collaboration and children's representation of the horizontal surface of liquid. *Journal of Applied Developmental Psychology, 19*, 571–592.

Gruber, H. E., & Wallace, D. B. (1999). The case study method and evolving systems approach for understanding uniquely creative people at work. In R. J. Sternberg (Ed.), *Handbook of Creativity* (pp. 93-115). New York: Cambridge University Press.

Guiller, J., Durndell, A., & Ross, A. (2008). Peer interaction and critical thinking: Face-to-face or online discussion? *Learning and Instruction, 18*, 187–200.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*(4), 395–429.

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science, 28*, 115–152.

Harasim, L. M. (1993). Global networks: Computers and international communication. Cambridge, MA: MIT Press.

Heckman, R., & Annabi, H. (2005). A content analytic comparison of learning processes in online and face-to-face case study discussions. *Journal of Computer Mediated Communication, 10*(2), article 7.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558.

Judge, G. G., Griffiths, W. E., Hill, R.C., Lutkepohl, H., & Lee, T. C. (1985). *The theory and practice of econometrics* (2nd ed.). Wiley: New York.

Kennedy, P. (2003). *A guide to econometrics.* Cambridge, MA: MIT Press.

Krippendorff, K. (2004). *Content analysis.* Thousand Oaks, CA: Sage.

Lindow, J. A., Wilkinson, L. S., & Peterson, P. L. (1985). Antecedents and consequence of verbal disagreements during small-group learning. *Journal of Educational Psychology, 77*, 658–667.

Nijstad, B. A., Diehl, M., & Stroebe, W. (2003). Cognitive stimulation and interference in idea generating groups. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration* (pp. 137–159). New York: Oxford University Press.

Pena-Shaff, J. B., & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education, 42*, 243–265.

Piaget, J. (1985). The equilibrium of cognitive structures. In T. Brown & K. L. Thampy (Eds.), *Trans*. Chicago: University of Chicago Press.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences, 2*, 235–276.

Schrire, S. (2004). Interaction and cognition in asynchronous computer conferencing. *Instructional Science, 32*, 475–502.

Swann, W. B., Jr., Kwan, V. S. Y., Polzer, J. T., & Milton, L. P. (2003). Fostering group identification and creativity in diverse groups. *Personality and Social Psychology Bulletin, 29*, 1396–1406.

Tallent-Runnels, M. K., Thomas, J. A., Lan, W. Y., Cooper, S., Ahern, T. C., Shaw, S. M., & Liu, X. (2006). Teaching courses online: A review of the research. *Review of Educational Research, 76*(1), 93–135.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge: Cambridge University Press.

# Concrete vs. Abstract Problem Formats: A *Dis*advantage of Prior Knowledge

Andrew F. Heckler, Department of Physics, Ohio State University Columbus, OH 43210
Email: heckler.6@osu.edu

**Abstract:** Three experiments examine the effects of varying the relative concreteness of physics word problems on student performance. Previous studies have found that concrete representations benefit performance for relatively simple problems, whereas abstract representations are beneficial for more complex problems. These findings are replicated in a physics context. More importantly, a significant disadvantage for concrete representations is identified for some questions. When a problem potentially elicits prior knowledge that is contrary to scientific knowledge, e.g. a scientific "misconception", it is found that the concrete representation invokes incorrect answers more frequently than abstract representations. In addition, an interaction is found between the final course grade of the student and the abstract vs. concrete problem format, with higher grade students performing disproportionally better on abstract problems. This is consistent with previous findings in reasoning and an explanation is provided that involves the cuing of familiar, automatic knowledge and skills vs. explicit deliberate processes.

## Introduction

One truism in education research is that student responses to a particular question depend on the how the question is represented. This idea can be applied in a productive way for the topic of problem solving: while it is clear that student performance on problem solving depends on the representation (or context) of the problem, one can gain a better understanding of the nature of student problem solving by determining *which* dimensions of representational variation matter and *how* the variations affects performance (e.g., Kotovsky, Hayes, & Simon, 1985; Collins & Ferguson, 1993, Zhang, 1997).

This paper builds on a series of previous studies that examine the effects of varying the dimension of concrete vs. abstract problem representations on student performance. These studies indicate that there is often a tradeoff between concrete and abstract representations. In particular, in cases where the problem complexity is substantial, abstract problem representations have been found to be beneficial for problem performance (Koedinger, Alibali, & Nathan, 2008), and have been found to facilitate transfer in problem solving performance (Goldstone & Son, 2005; Kaminsky, Sloutsky & Heckler, 2008).

In contrast, concrete representations can be beneficial because, for example, previous knowledge can facilitate situated reasoning that allows for the use of familiar, intuitive strategies practiced frequently (Kintsch & Greeno, 1985; Carraher, Carraher, & Schliemann, 1987; Nathan et al., 1992; Pollard & Evans, 1987) and provides redundancy that is helpful for avoiding and detecting errors (e.g. Koedinger, Nathan, 2004; Koedinger, Alibali, & Nathan, 2008).

We consider two points regarding abstract and concrete problem representations. First, while grounded or concrete problem representations may produce problem-solving benefits via the cuing of helpful prior knowledge of a specific situation, it may also be the case that the grounded representation can cue knowledge that drives the solver toward an *incorrect* solution path. Potential examples of this are the well known findings of patterned and "incorrect" student answering to simple conceptual questions about physical phenomena (e.g., Pfundt & Duit, 2000). Some have interpreted these findings as evidence of coherent "scientific misconceptions", while others have explained it in terms of the cuing of "knowledge in pieces" or "resources" (diSessa, 1988; Hammer, 2000). In either case, the patterned incorrect answers are a result of prior knowledge of some type.

The second point to consider is that individual differences can elucidate important insights into the nature of the effect of representational variation. For example, in reasoning studies involving the Wason card task, Stanovich and West (1998) provide evidence that both high and low SAT (test score) participants can easily solve the task in some concrete, familiar representations, but in the generic form the task is more difficult and is solved much more frequently by high SAT participants. They argue that concrete, familiar tasks tend to invoke practiced and automatic knowledge that is independent of ability, and abstract tasks require higher order, more deliberate thinking skills.

Therefore in this paper we address two questions. First, if prior knowledge can in some cases include "incorrect" scientific knowledge, then in these cases will concrete, familiar representations tend to cue this incorrect knowledge more frequently than abstract and less familiar representations? Put more plainly, will students display misconceptions more frequently in concrete problems than abstract problems? Second, will there be a significant interaction between ability level of the student and the representational format? A disproportionate increase in performance for high ability on abstract questions students would lend more support for the idea that concrete representations cue familiar and automatic solution methods that are common to all students.

## Experiment 1

The goal of the first experiment is to replicate previous studies which demonstrate that concrete, more grounded problems are more successfully solved than more abstract, generic problems if the problems are sufficiently simple and familiar (e.g., Koedinger et al., 2004; 2008). A simple physics problem is posed in two different contexts: abstract and generic vs. concrete and familiar. In addition, the problem is constructed to include a physical situation which is known to elicit misconceptions.

### Participants

The participants were 170 university students enrolled in the first quarter of a calculus-based introductory physics course at a large research university. They completed this assignment for participation credit as part of their total course grade. Participants were randomly placed in one of two conditions, the abstract (n=92) and concrete (n=78) contexts conditions.

### Materials, Design, and Procedure

Participants were given one of the two physics problems shown in Table 1. They were given 15-20 minutes to complete the task at the beginning of a physics lab class in a proctored environment, and were given credit for participation. The problems are simple mechanics problems that would be typically found in the back of a textbook or on an exam, and are given routinely in homework assignments. The problem is typically solved by determining the two unknowns in the problem (friction and acceleration) via solutions to (a minimum of) two simple equations, one involving the definition of friction force ($F_{friction} = \mu F_{\text{normal}} = \mu mg$) and Newton's second law ($F_{\text{net}} = F_{applied} - F_{friction} = ma$).

While it may not be obvious at first glance, even to the expert, the physical situation described by the problem is somewhat counterintuitive. In particular, the net force resulting from the two horizontal forces on the object (applied force and friction force) is in the direction opposite the motion. That is, the applied force is less than the friction force. Nonetheless, the object can still be moving, just slowing down. It is well-known that many students have difficulty with the idea that the net force can be opposite the direction of motion (Viennot, 1979; Clement, 1982).

Table 1. Question Set 1: isomorphic abstract and concrete versions of a force-motion show-work problem.

| Abstract version | Concrete version |
|---|---|
| A force F is exerted on an object in the positive direction. The object is sliding on a surface in the positive direction toward a mark on the surface. At the very moment the object passes the mark, the force F is 80N. The mass of the object is 32Kg and the coefficient of friction between the object and the surface is $\mu_k$=0.3.<br>What is the magnitude and direction of the acceleration of the object at the moment it passes the mark? | In front of a house, a boy is pulling on a large toy rocking-horse in the direction towards the driveway. The rocking-horse is on the sidewalk sliding toward the driveway. At the very moment the rocking-horse crosses onto the driveway, the boy is pulling it with a force of 80N. The mass of the large toy rocking-horse is 32Kg and the coefficient of friction between the toy and the concrete is $\mu_k$=0.3.<br>What is the magnitude and direction of the acceleration of the toy rocking-horse at the moment it crosses onto the driveway? |

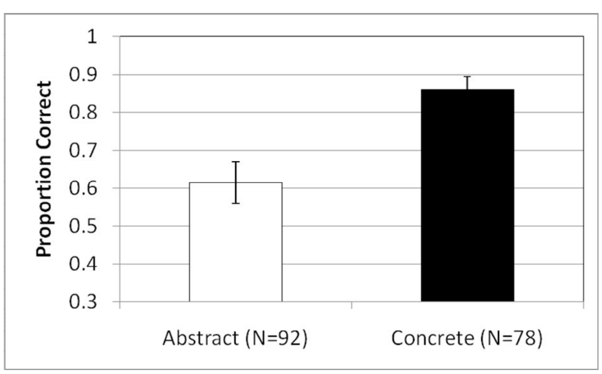



Figure 1. Student performance on Question 1, separated by condition. Error bars are 1 SEM.

## Results and Discussion

As indicated in Figure 1, students were significantly more successful in obtaining a correct solution in the concrete condition (86%) than the abstract condition (62%), $\chi^2(1) = 13.2$, $p < .0001$, effect size $d = .6$. These results are consistent with previous findings that concrete, grounded context results in better student performance compared to the abstract context Further analysis of the solutions supported a previously described explanation for this difference (e.g. Koedinger et al. 2004), namely that the abstract condition inhibits the use of familiar knowledge that may help in understanding and solving the problem. In the abstract condition, the most common errors involved issues of not properly understanding the physical situation described in the problem text. For example in the abstract condition a 10/92 students confused (i.e., switched) the values of the two of the forces provided in the problem (the applied force and the weight, via mass). In addition, 5/92 students in the abstract condition ignored friction. In contrast, none of the students in the concrete condition made these errors; the most common error in the concrete condition was a simple algebraic error. Therefore, these results suggest that students in the abstract condition were less able to understand the physical details of the problem, and this lack of physical understanding could in turn inhibit them in their solution to the problem.

However, responses from a small number of students suggest that in some cases the concrete context can interfere with success in solving the problem as well. In particular, the common conceptual error that an object must be moving in the direction of the net applied force was made by a small number of students (4/78) in the concrete condition. For example, one of these students wrote, "Because net force is less than zero, static friction was not overcome and acceleration is zero". In contrast, *none* (0/92) of the students in the abstract condition made this error. However, the elicitation of this misconception may have been minimized in the particular concrete context given, as the friction force (large toy on concrete) is fairly salient compared to the force in the direction of motion (via a boy). This may partially explain why only a small number of students displayed the misconception.

## Experiment 2

Experiment 1 replicated the finding that the concrete condition provides benefits to problem solving. However, there was also an indication that more concrete contexts can cue misconceptions more frequently. The purpose of Experiment 2 is to determine more directly whether a concrete context cues a misconception more frequently than an abstract context. We elicit the same misconception as in Experiment 1, namely the student belief that the direction of the net force on an object and its velocity must be in the same direction.

### Participants

Two hundred and fifty-two participants were enrolled in the first quarter of a calculus-based introductory physics series at a large research university. They completed this assignment for participation credit as part of their total course grade.

Table 2. Question Set 2: isomorphic abstract and concrete versions of a force-motion conceptual question.

| Abstract version | Concrete version |
| --- | --- |
| At a particular instant of time, there are several forces acting on an object in both the positive and negative direction, but the forces in the negative direction (to the left) are greater. Which statement best describes the motion of the object at this instant?<br>a) it is moving to the right<br>b) it is moving to the left<br>c) it is not moving<br>d) both a and b are possible<br>e) both a and c are possible<br>f) a, b, and c are possible | A force sensor is attached inside a soccer ball that is used during a match. The force sensor measures the forces acting on the ball. At a randomly chosen instant during the game, the sensor detects that there is only one horizontal force on the ball, and that force is directed toward the home-team goal. Which statement best describes the motion of the ball at this instant?<br>a) the ball is moving toward the home team goal<br>b) the ball is moving away from the home team goal<br>c) the ball is not moving<br>d) both a and b are possible<br>e) both a and c are possible<br>f) a, b, and c are possible |

Note: For both questions, answer *f* is the correct answer.

### Materials design and Procedure

Experiment 2 is a within-subject design. Participants completed the multiple choice questions shown in Table 2 in a quiet, proctored room. These two questions were part of a larger set of multiple choice questions and the design was counter-balanced for question order (no reliable differences in order were found). The validity and reliability of the multiple choice questions was verified through extensive testing and interviews (Rosenblatt, Sayre & Heckler, 2008). These two questions are designed to elicit student beliefs about the relationship between the direction of net force on an object and the direction of its velocity[1].

The final course grades of the students were also collected. The student were then categorized as the High grade students (n =123) and Low grade students (n = 119), indicating whether they were in the top half or bottom half of the class as measured by the final course grade.

## Results and Discussion

While both questions were difficult overall, students answered the Abstract question correctly more often (20%) than the Concrete question (9%), McNemar's test $p < .001$, effect size $d = 0.32$. Figure 2 shows the proportion of students choosing the correct answer choice separated out by High and Low Grade students. Inspection of Figure 2 reveals that High and Low grade students perform similarly (poorly) on the Concrete question and the High Grade students perform significantly better on the Abstract question than the low grade student. A repeated measures analysis, with question type as the repeated measure and grade rank as the between student factor, reveals significant main effects of question type ($F(1) = 18.4$, $p < .001$), and rank ($F(1) = 4.5$, $p = .035$), and there is a reliable interaction between question type and grade rank, $F(1) = 8.0$, $p = .005$. This interaction is mainly due to the High Grade students choosing the correct answer more often for the abstract question (26%) than for the concrete question (10%), McNemar's test, $p < .001$. There is no significant difference in choosing the correct response between the concrete (12%) and abstract (8%) question for Low grade students, McNemar's test $p = .5$.

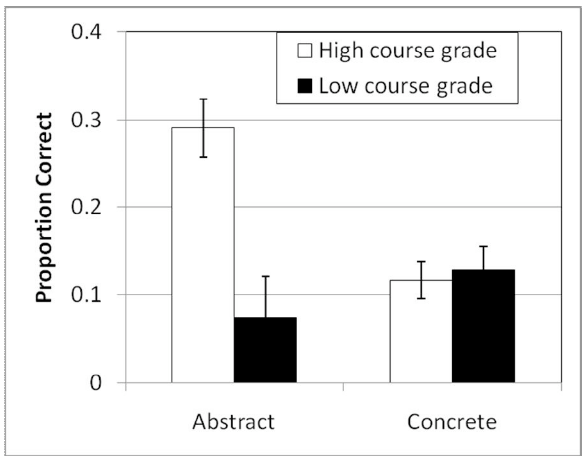



Figure 2: Student performance on Question set 2, separated by High course grade students (n = 123) and Low course grade students (n =119). Error bars are 1 SEM.

## Experiment 3

The purpose of this experiment is two-fold, first it is to replicate the finding that while concrete context help with simple problems, more complicated problems may have increased performance in more abstract problem representations (Koedinger et al., 2008). Therefore, we pose a slightly more complicated problem (involving three rather than two equations and unknowns). The second purpose is somewhat exploratory. Experiments 1 and 2 indicate that everyday knowledge may in some cases hinder problem solving because it may misguide the solver to an incorrect solution path. In Experiment 3, we pose a simple kinematics problem that lends itself to using everyday knowledge, but may also present cues in the concrete context that lead to incorrect assumptions.

### Participants

The participants were 68 university students enrolled in the first quarter of a calculus-based introductory physics course at a large research university. They completed this assignment for participation credit as part of their total course grade. Participants were randomly placed in one of two conditions, the abstract (n=32) and concrete (n=36) contexts conditions.

### Materials design and Procedure

Participants were given one of the two physics problems shown in Table 3. They were given 15-20 minutes to complete the task in a quiet, proctored room. The problems are simple kinematics problems that would be typically found in the back of a textbook or on an exam, and are given routinely in homework assignments. The problems are typically solved by determining three unknowns, the distance that each person travelled and the acceleration of one of them. This involved using three equations: the total distance $d_{tot} = d_1 + d_2$, the kinematic equation for the constant velocity person $d_1 = v_1 t$, and the kinematic equation for the constant acceleration person $d_2 = a_2t^2/2$. Therefore this problem is slightly more complicated than the problem in Experiment 1, which only needed two equations and two unknowns.

Table 3. Question Set 3: isomorphic abstract and concrete versions of a kinematics show-work problem.

| Abstract version | Concrete version |
|---|---|
| Two objects are 60 kilometers apart and move toward each other such that they will meet at a point in-between the original starting positions. The first object travels in the positive direction toward the meeting point at a constant speed of 18 meters/second. The second object begins at zero speed and constantly accelerates in the negative direction toward the meeting point. If both objects meet in exactly 30 minutes, what was the acceleration of the second object? | Two friends, Kevin and Claire, live 60 kilometers apart and decide to meet in a town called "Tristville" in between them. Claire drives eastward toward Tristville at a constant speed of 18 meters/second. Kevin begins at zero speed and constantly accelerates westward toward Tristville. If they both reach Tristville in exactly one-half hour, what was Kevin's acceleration? |

## Results and Discussion

There are three important results from Experiment 3. First, unlike Experiment 1, students were equally successful in obtaining a correct solution in the concrete condition (47%) than the abstract condition (53%), $\chi^2(1) = .2, p = .6$. Second, Figure 3 shows the proportion of students obtaining the correct solution, separated out by High and Low Grade students. Inspection of Figure 3 reveals a pattern similar to the results in Experiment 2: High and Low grade students perform similarly on the Concrete question and the High Grade students perform significantly better on the abstract question than the Low Grade students. A two-way Generalized Linear Model (question type × grade rank) with a binary response analysis reveals a significant main effects of rank (Wald $\chi^2(1) = 9.9, p = .002$), and there is a reliable interaction between question type and grade rank, Wald $\chi^2(1) = 4.1$, $p = .04$. This interaction is mainly due to the High Grade students choosing the correct answer more often of the abstract question (87%) than for the concrete question (57%). While there may be a trend for students choosing the correct response in the concrete condition (41%) compared to the abstract condition (24%) for Low Grade students, the difference is not statistically reliable.

The third result comes from a closer examination of the solution methods which reveals an unanticipated and fairly frequent misunderstanding of the problem. In particular, 28% (10/36) of the students in the concrete condition made the same mistake of explicitly assuming that the two friends in the problem met *halfway* when in fact it is only stated in the problem that they met "in between" the starting points. None (0/32) of the students in the abstract condition made this erroneous assumption. Students' remarks in debriefing revealed that the most plausible explanation for this assumption is that the students in the concrete condition assumed that "friends meet halfway". In the abstract condition, the "friends" are replaced with "objects", and none of the students appeared to be cued to assume that the objects met halfway. Therefore, this is a case in which the concrete context appears to be interfering with the correct solution. It is important to note however, that the fraction of students in the concrete condition making this assumption did not depend on grade rank $\chi^2(1) = .8$, $p = .4$. Therefore, if this "misconception" (which could also be viewed as simply an incorrect assumption) is responsible for the difference in score between the abstract and concrete conditions for the High Grade students, it remains to be explained why the Low Grade students did not have a similar gain in score. It may be that the low grade students are such poor performers that switching from concrete to abstract simply involves switching from one incorrect method to another incorrect method.

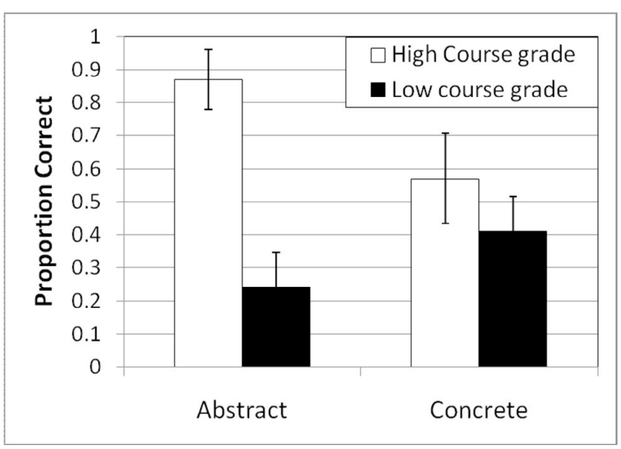



Figure 3: Student performance on Question Set 3. Error bars are 1 SEM.

## General Discussion

There are four main results from this study. First, the advantage of concrete problem representation found in Koedinger et al. (2004, 2008) was replicated for a relatively simple physics problem. This experiment was different from the previous studies in that both of the representations were in "story" format. Nonetheless, the nature of the advantage in the concrete representation, namely better comprehension of the problem appears to be consistent with these previous studies.

The second finding was a significant interaction between abstract vs. concrete problem representation and student course grade for a simple conceptual question that is known to strongly elicit a well-known scientific misconception. In particular, the concrete, familiar representation was likely to elicit the misconception at a similarly high level for both High and Low Grade students; however, in the abstract representation, the High Grade students performed significantly better than the Low Grade students. This result could be interpreted as support for the idea that incorrect answering in the concrete representation is due to cueing of everyday, familiar knowledge common to all students while the abstract representation is less likely to cue such knowledge. The High Grade students may answer correctly more often in the abstract context because they may tend to use explicit reasoning skills rather than on relying on automatic, prior knowledge. Another possible explanation is that both High and Low Grade students began the course at floor on these questions, and the High Grade students have learned the abstract representation simply because the course may have presented more abstract than concrete examples, thus they would perform better on the abstract questions. In contrast, the Low Grade rank students may have simply not learned the material yet, thus they score low on both abstract and concrete questions. Understanding the distinction between these two possibilities is worth further investigation, and it is possible that both may be at work.

The third finding was a lack of difference in class averaged performance of the concrete and abstract representations for a problem that was slightly more complex (as measured by number of equations and unknowns to solve) than the problem in Experiment 1. This is somewhat consistent with the findings of Koedinger et al. (2008) in that they found an increase in advantage of the abstract representation when the problem difficulty was increased.

The fourth finding is similar to the second, but for a quantitative problem rather than a conceptual question. In particular, High and Low Grade students perform similarly in the concrete representation, but High Grade students perform significantly better in the abstract representation. A similar proportion of High and Low Grade students displayed an incorrect assumption in the concrete representation that was apparently caused by the familiar nature of the representation, and no students displayed this assumption in the abstract condition. Once again it is not clear what causes the difference in performance between the High and Low Grade students. The two explanations given above for the second finding remain viable. The concrete representation may be cueing familiar automatic problem solving processes and the abstract may be cuing (in High Grade students) more explicit, deliberative processes. On the other hand, it may be the case that (effectively) only relatively abstract examples are practiced in the course and only the High Grade students have learned them. In any case, it is clear that the concrete representation is cuing knowledge that inhibits the solution process, at least in High Grade students.

In sum, while some advantages of concrete representation have been replicated, this study identifies an important disadvantage for concrete representations: they can cue familiar knowledge that is contrary to scientific knowledge and this in turn can lead solvers to incorrect solution paths. This raises some practical issues about the relative usefulness of concrete and abstract questions as a diagnostic instrument. From a psychometric perspective, the abstract representation would be preferable over the concrete representation because the former is more successful at discriminating between students, at least as measured by final course grade. However, from a pedagogical perspective, the results from this study suggest that both high and low level students are susceptible to misconceptions cued by problems posed in the concrete representation format. Therefore, concrete problem representations are important for determining whether both High and Low Grade students have overcome the relevant scientific misconceptions. Once again, these results highlight the important tradeoffs between problems with concrete and abstract representations.

## Endnotes

(1) Note that the two questions are not exactly isomorphic: the abstract question mentions forces in two directions and the concrete question mentions only one force in one direction. However, in our previous work in constructing items of this type we found that this difference does not reliably or consistently change student answering patterns.

## References

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1987). Written and oral mathematics. *Journal for Research in Mathematics Education, 18,* 83–97.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50,* 66-71.

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist, 28*, 25–42.
diSessa, A.A., (1988). Knowledge in pieces. In G. Forman & P.B. Pufall (Eds.) *Constructivism in the Computer Age* (pp. 49-70). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences, 14*, 69–110.
Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics, 68*, S59.
Kintsch, W., and Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92,* 109–129.
Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-Offs Between Grounded and Abstract Representations: Evidence From Algebra Problem Solving. *Cognitive Science*, *32*, 366-397.
Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: effects of representations on quantitative reasoning. *Journal of the Learning Sciences*, *13*, 129-164.
Kotovsky, K. , Hayes, J.R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17(2):248–294.
McCloskey, M., Caramazza, A., & Green, B., (1980). Curvilinear motion in the absence of External forces: Naïve Beliefs about the motion of objects. *Science*, 210, 1139-1141.
Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra word problem comprehension and its implications for the design of computer learning environments. *Cognition and Instruction, 9*(4), 329–389.
Pfundt, H., Duit, R. (2000). *Bibliogaphy: Students' Alternative Frameworks and Science Education* (5th Ed.). Kiel, Germany: Institute for Education Science.
Pollard, P., & Evans, J.St.B.T. (1987). Content and context effects in reasoning. *American Journal of Psychology, 100*, 41–60.
Rosenblatt, R., Sayre, E. C., & Heckler, A.F. (2008). "Toward a comprehensive picture of student understanding of force, velocity, and acceleration," in C. Henderson, M. Sabella, & L. Hsu (Eds.) *Proceedings of 2008 Physics Education Research Conference*. Melville, New York: AIP Conference Proceedings.
Sloutsky, V. M., Kaminski, J. A. & Heckler, A. F. (2005). The advantage of simple symbols for learning and transfer. *Psychonomic Bulletin & Review, 12*, 508–513.
Stanovich, K.E. and West, R.F. (1998) Cognitive ability and variation in selection task performance. *Thinking and Reasoning* 4, 193–230.
Viennot, L. (1979). Spontaneous reasoning in elementary dynamics, *Eur. J. Sci Educ*. **1**, 205-221.
Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*, 179–217.

## Acknowledgments

This research is supported by the Institute of Education Sciences, U.S. Department of Education, through grant #R305H050125. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

# Tension resolution as pattern for practice transformation in interdisciplinary teamwork in professional development

Patrick Sins
Research Centre Learning in Interaction, Faculty of Social and Behavioural Sciences,
Utrecht University, PO Box 80410, 3508 TC, Utrecht, The Netherlands
p.h.m.sins@uu.nl

**Abstract**: A central interest in developing professionalism resides in the potential for professionals to learn from and with one another in ways that support transformations of their knowledge practices. However, negotiation between multiple perspectives, interests, practices and traditions intertwines cognitive-epistemic with socio-relational and affective aspects, which may lead to tension and conflict. While tension can disable learning, this paper argues that identifying these tensions should be viewed as a significant source for change and development. The paper reports on a teacher-researcher collaboration at a secondary school which focuses on the co-design of a learning module. It is shown that the identification of tensions during meetings helped participants to focus their efforts on the root causes of problems, which led to a reconceptualisation of the current work practices. This subsequently helped team members to deviate from established norms and improve their practices.

## Introduction

Rapid changes in current networked society present new challenges to human competence and flexibility. Productive participation in knowledge-intensive work requires that individuals, their professional communities, and their organizations develop new practices, advance their knowledge and their understanding as well as produce innovations. This is reflected in developments in professional communities wherein work is increasingly focused on the deliberate advancement of knowledge rather than on the mere production of material objects (Bereiter, 2002; Paavola & Hakkarainen, 2005). In order to cope with the cognitive, social, and motivational challenges of knowledge-based society, tools and pedagogical methods are needed that open up opportunities for enduring and sustained transformation of professionals' knowledge practices.

Based on the works of Engeström (1987), Schatzki (2002) and Reckwitz (2002) we define a knowledge practice as follows: *a social-historically created and shared behavioural pattern consisting of an interconnected and inseparable array of recurrent activities, conventions, rules and norms that play part in the creation of knowledge artefacts.* According to this conceptualization, practices are characterized by their social nature, which means that practices are shaped by and evolve within a knowledge community, ultimately becoming part of the its identity. In addition, the concept of knowledge practice entails stability as well as change. Stability is reflected as routines, procedures, conventions, underlying beliefs and values, epistemological conceptualizations and the set of available tools. At the same time, practices are open to change in that each activity based on this practice is adapted in response to changing contexts and particular circumstances.

The rationale for practice transformations is that newly developed practices and tools aim to overcome the tensions a particular knowledge community identified during certain events in their knowledge work. Practice transformations involve fundamental changes in views, beliefs, ideas and ways of working with knowledge that fulfil a certain need that is relevant for a particular professional community's knowledge work. These transformations lead towards historically new types of practices based on collaborative, tool-mediated knowledge production that takes place as long-term, sustained processes and which ultimately lead to a reconceptualization of the object and motive of the community's knowledge practices to embrace a more diverse horizon of possibilities than in the previous practice.

However, practices are difficult to change, not only because this would imply a negative evaluation of previous socially grounded practices, but also because such transformation involves a period of disorientation while old practices are gradually unlearned and new practices are gradually developed (Eraut, 2004). During this period professionals feel like novices, but without having the excuses or discounts on performance normally assigned to novices. The pain of transformation lies in the loss of control over one's practice when one's tacit knowledge ceases to provide the necessary support; and the emotional dimension is also of considerable importance. In addition, Little (1990) reports that professionals view transforming practices as involving high transactional costs to participatory work in time. According to Argyris and Schön (1978) the central problem for most professionals is that they are intellectually and emotionally committed to espoused theories which describe the world as they would like it to be, but which do not necessarily accurately describe their own activities and constrain possibilities for transforming their practices. Moreover, practices are similar to physical infrastructures in a sense that when everything is working well one does not pay attention to them. Consequently, professionals rely on them even they are not fully aware what constitutes them. Additionally, the practices have been evolved during a long period. According to these authors, these problems can only be solved when professionals step outside their taken-for-granted world and espoused theories to actively search for genuine feedback which challenge the outcomes of their activities.

Although practices tend to remain stable and to reproduce socially shared knowledge, this stability is altered as internal or external disturbances produce various forms of tensions within activity systems (Engeström, 1987, 2007; Hakkarainen, Palonen, Paavola, & Lehtinen, 2004). The notion of developmental tensions as a driving force of change and development is drawn from the premises of cultural-historical activity theory (CHAT) (Engeström, 1987; 1999). CHAT draws upon the works of Heidegger (1962), Dewey (1968) and Leont' ev (1981) converging on the role of tensions and their resolution as a means of revealing the nature of the world around us. In these socio-historical approaches to learning and knowledge, tensions are conceptualized as the conditions that open up opportunities for creative efforts in activity and communication and are as such the driving force behind innovative knowledge practices.

Changes in practices are due to external as well as internal disturbances. According to Engeström (1999a; 1999b), identification of tensions in an activity system helps learners to focus their efforts on the root causes of problems. Barab, Barnett, Yamagata-Lynch, Squire, & Keating (2002) corroborate on this in arguing that tensions and their resolution helps to identify the dynamic forces of change and comprise an important constituent and starting point for investigating such processes. The mechanism behind tensions resolution is that as the tensions of a particular activity system are aggravated, some individual participants begin to question and deviate from its established norms. Learning is accomplished when tensions lead to a reconceptualization of the object and motive of a particular activity to embrace a more diverse horizon of possibilities than in the previous activity. While practices might change when new tools become available or circumstances and contexts shift, they can also be deliberately altered by those carrying out the activities when they invent new strategies of working or activities. In addition, members of a community may experience difficulties in constructing a connection between the goals or their actions and the object and motive of the collective activity, which may give rise to tensions.

For instance, knowledge communities that involve cross boundary practices between different domains of knowledge embody an epistemic collision of activity systems which generates disturbances and conflicts. This involves that in collaborative contexts, negotiation between multiple perspectives, interests, practices and traditions intertwines cognitive-epistemic with socio-relational and affective aspects, which may lead to tension and conflict. As practices inevitably change when concrete activities are carried out, the development and transformations of practices proves to be an ongoing and contingent process. Due to this dynamic nature of practices, the design of a new object might overcome shortcomings of former existing objects, but when they are employed as tools this will restructure the nature of the current practices and will in this manner create new opportunities but also new problems and challenges. According to this line of reasoning, learning or development of practices occurs when members attempt to negotiate about how to adapt their knowledge practices as a results of these disruptions feeding into the shared object.

In general, what is important here is that a group of professionals has to capitalize on the multiple perspectives of its members, that the group creates ownership and commitment towards achieving the common objectives and that members build up on each other willing to learn from tensions. The way to overcome these tensions is to negotiate possible solutions or design hypotheses for new practices or new tools. Eventually these new practices and tools lead to new tensions which leads to an iterative sideways movement. Preconditions for groups to be able to solve tensions to improve their practices are: 1) proactive and continuous creation of common ground, 2) sharing tensions in considerate ways and 3) supporting and recognizing agency (Matusov, 2001). However, what is still lacking in socio-historical accounts and studies of tension resolution as conditions for practice transformations to occur, is a description of patterns of tension resolution that lead to productive changes and development. Moreover, the investigation of trajectories of tension resolution affecting the emergence of new knowledge practices located at several timescales and places has rarely been touched upon in empirical endeavours.

In the present paper, we will present the data of a study in the field of a school-university partnership at a secondary school which shows a particular patterns in the cognitive-epistemic consequences and resolution of tensions. This study focused on the investigation of the practice changes of an interdisciplinary team which focused on the redesign and implementation of a learning module. Meetings of this team served to reflect upon and analyze teachers' prevailing practices, as well as envision and implement changes to the module.

## The Setting and Knowledge Practices of the Investigated Knowledge Community

This study focuses on describing the nature and scope of transformations in teachers' coaching practices that followed from resolutions of developmental tensions that arose in an university-school partnership. This partnership involved the 2-year lasting collaboration at UniC, a secondary school in Utrecht in The Netherlands that was aimed at the redesign of a course module based on knowledge creation principles (Paavola & Hakkarainen, 2005). One central feature of the knowledge creation approach that was taken up in the design was the concept of *mediation* which means that students' activities are directed towards the collaborative creation and advancement of shared knowledge objects (e.g., documentaries, research reports or instructional material) mediated by specific supporting technological or conceptual tools.

At UniC, students are coached towards the national school exam, complementing the focus on knowledge acquisition by stressing development of competencies, skills and personal development. By clever organization of compulsory learning materials students are enabled to develop their own talents and interests in a course module in which they plan and perform projects within or outside of the school context. This means that every week in the curriculum a half day is reserved for these projects for periods each of which last eight weeks. The school supports the students and offers possibilities to carry out their projects.

Within this pedagogical context, teachers' coaching practices traditionally focus on the development of courses and assignments providing guidance to students' self-directed learning process. UniC expressed the aspiration to challenge their students more towards meaningful learning. In addition, the teachers' expressed that their role during these projects was unclear and that they needed more scaffolds to structure their coaching. Therefore, a multi-disciplinary team consisting of educational researchers, teachers, students, dean, process coordinator and pedagogical experts was set up to flesh out the design based on knowledge creation principles which matched UniC's general pedagogical approach and objectives. The collaborative design, implementation and testing of the new course module implied that high demands were placed on coaching practices of the teachers which provide a platform for tensions to arise. For instance, they had to: (a) comprehend the theoretical principles behind the knowledge creation metaphor, (b) apply these principles in their practice, and (c) reflect on their role as a teacher and transform their practices accordingly to scaffold students' knowledge creation processes.

## Method

Multiple, intertwined methodological approaches and various approaches to data collection and analysis were combined to elaborate dynamics of incremental changes which reflect practice transformations resulting from patterns of tension resolution. The study involved ethnographic analyses with participatory observation, a developmental intervention approach, interviews and event sampling to follow the processes towards new practices. Our analyses took tensions in partners' activities as a point of departure. We look for episodes in the material that express problems and materializes as developmental tensions.

Then we can investigate discursive activities between partners (micro level), elaborate on episodes of tension resolutions over time (meso-level) and examine how patterns of tension resolution relate to transformations of practices at the level of trajectories (macro-level). This gives to transformation processes as they play as professionals deliberately extending shared knowledge and mobilize knowledge types embedded in the social practice in their problem solving. For framing the analysis we developed the following approach:

1. Description of the knowledge practices of the investigated professional community;
2. Description of the nature of tensions that occur in these practices. Explore resolutions to the tensions, characterize tension-resolutions and analyze how these contribute to transformation towards new practice over time;
3. Examination of practice transformation including emphasizing the vertical (socialization and internalization) and/or horizontal dimension (boundary crossing and externalization).

The cases in this cluster examine the nature of these tensions that provide an important starting point to explore change and development of knowledge practices. To assess the knowledge practices of the actors involved both studies, we combined data collected from different instruments, namely:

- Material artefacts, such as reports, concept maps, and written comments;
- Pre- and post questionnaires;
- Semi-structured interviews;
- Transcribed recordings and minutes of the meetings.

These instruments were mainly designed to capture critical events during the meetings and to discern how these events are echoed in the ways professionals adapted their practices. Critical events were conceptualized as articulations of developmental tensions and were posteriori checked with the investigated knowledge communities.

## Results

Most tensions were observable on the boundary of the intersecting activity systems showing how team members balanced institutionalized or traditional and newly developed practices. One source for tensions involved the specific organization of teachers' coaching practices to be more in line with the new pedagogical approach and at the same time foster students' knowledge construction processes. The following excerpt exemplifies this tension during an interview with one of the participating teachers:

| | | |
|---|---|---|
| 1 | Teacher3 | I see that an increasing amount of student groups do not have a clear view of what they are doing, that is what I am afraid of, unfortunately |
| 2 | Researcher1 | How do you coach these students then? |
| 3 | T3 | Well, you cannot just leave them, than this would lead to chaos. […] You can divide tasks in the group and think of who is going to do what, but then I would be too directive and I am not sure whether that should be our intention, so therefore I give them more freedom […] |
| 4 | R1 | […] Well you mean that you are still in search of what is expected of you as a teacher. What do you need in your coaching? |
| 5 | T3 | First I need to know more about knowledge creation, what the idea and what the pillars are, so I can eventually adapt my coaching to that […] normally I am very clear in my teaching, but in this pilot it seems that you have to discover what the best ways of coaching are |

*Interview Teacher 3; December 2006*

Teacher3 expresses his concern about his impression that students have not been successful in organizing and structuring their work. When prompted for his ways to cope with this impression in his coaching, he states that he would like to be more directive, saying " .. you can divide tasks in the group" (passage 3). He identifies a tension with what he interprets as the coaching practice which would comply with principles of knowledge creation" but then I would be too directive and I am not sure whether that should be our intention, so therefore I give them more freedom" (passage 3). Eventually, for him to overcome this dilemma, more guidelines for coaching knowledge construction are needed.

Thus, this episode points to a tension between what can be interpreted as top-down instruction at one hand and social distancing at the other. Teacher3 was reluctant to interfere with students' activities too much, since this would be in conflict with his emergent perspective about what is important when fostering students' knowledge construction. Still, this coaching practice added to the problem of students, who reported that they sometimes experienced their teachers' approach as being disruptive in the context of their activities:

| | |
|---|---|
| Researcher3 | What do you think of your teachers' coaching? |
| Student1 | Well, sometimes teachers ask us just too often what we are doing and what our end product will be; what we want to achieve. But in most cases, we had explained that seven times already and they still want us to explain it even further, while we even do not yet know how far we can go, that is quite annoying |

*Interview Student group 4; December 2006*

In this episode student1 reflects on their teacher's approach, mainly involves asking explanatory questions about the nature and status of the group's activities. This is corroborating on what teacher3 stated in the previous episode. Although teacher3 adopted this coaching practice as his interpretation of what is needed to foster students' knowledge construction processes, this caused a conflict with the observed and experienced processes of teachers and students. This point was explicated during a subsequent plenary meeting of the design group:

| | | |
|---|---|---|
| 1 | Teacher3 | […] Well, it seems our students do not have a clear idea of what they have to do |
| 2 | Dean | Students have to know what the assignment entails […] |
| 3 | Ped. Expert | […] So I would suggest that the teachers can focus on helping to students create these structures. […] |
| 4 | Researcher3 | Students could concretize their ideas in a plan |
| 5 | T2 | So I would like to coach them to make it clearer like what the object is and its requirements. So far, we have maybe been too reserved. |
| 6 | PE | That is very important, and then those group members will follow their own work structure. […] |
| 7 | R2 | Still, it is not a bad thing that it going like this, if they first muddle a little, […] |
| 8 | PE | But you shouldn't let that continue too long |
| 9 | T1 | But, what you see now. That we should give a little more |

| | | |
|---|---|---|
| | | structure |
| 10 | T2 | We can ask students to make a so-called Tabasco planning which they are already used to construct. In this planning they have to specify the activities they are going to perform, what the end objectives are. This can serve as a tool for teachers to monitor students' progress without being too directive or strict […] |
| 11 | Process coordinator | And you could revise this planning, which makes leaves it more open for students […] |
| 12 | T2 | Apparently that is needed |
| 13 | R3 | So, it is our observation that that is needed, yes. |
| 14 | T2 | Well, that is clear by now. This shows that a good start is necessary. There has to be concrete object and once that it is there, it will go well |

*Protocol meeting design team; January 2007*

This episode shows the identification of the issue that is at hand ".. it seems that our students do not have a clear idea of what they have to do" (passage 1), leading to an expression and framing of the tension between teachers' current approach ".. So I would like to coach them to make it clearer like what the object is and its requirements. So far, we have maybe been too reserved" (passage 3), and the more directive approach ".. But, what you see now. That we should give a little more structure" (passage 9). The suggestion that is put forward to overcome this tension is to synthesize both perspectives in a concrete manner, ".. We can ask students to make a so-called Tabasco planning which they are already used to construct" (passage 10). This is accepted and taken up by the others ".. apparently that is needed" (passage 12).

Another issue related to teacher's 3 concern regarding the lack of guidelines for coaching students' knowledge construction, and the contribution and roles of the members of the design team in the coaching practices:

| | | |
|---|---|---|
| 1 | T3 | Nevertheless, it is important get more assistance during work sessions because now we're only with the three of us,. that is my first concern |
| 2 | D | It should be fixed then, we need teachers for this class […] |
| 3 | T1 | Sometimes you [Researchers] are a little blunt It is not criticism but I noticed that you have you own agenda You don't really help us coach, we just have to take care of it. In my opinion that is not really being an actor! |
| 4 | R1 | Well, the idea was that we didn't want to participate as a teacher because we don't have that expertise though we are here to provide you with some advice and answer your questions, if you have any |
| 5 | T1 | […] I am teaching the knowledge creation project on my own which is not an ideal situation, I just want you to think with me. Clearly we don't expect you to teach […] |
| 7 | R1 | Well, I believe that is a good thing to hear, I am glad that this came forward |
| 8 | PE | The researchers are used to stay in the background to be able to observe the process as objectively as possible |
| 9 | T2 | There is a big culture difference because we are used that everyone is involved You are think as observers |
| 10 | PC | It is a type of participation when you are observing […] |
| 11 | T2 | You could divide one group into two groups so that T1 has to coach his own groups but that T4 and T1 meet each other during class to discuss any problems or to ask each other for advice |
| 12 | D | You can then also ask researchers for feedback during coaching […] |
| 13 | R1 | Yes, that would be perfect [everybody agrees] |

*Protocol meeting co-design team; January 2007*

This episode shows a tension between the work traditions of researchers at the one hand and teachers at the other. The issue that was expressed "..it is important get more assistance during work sessions because now we're only with the three of us" (passage 1), opened up for explicating the underlying conflict between the perspectives of researchers and teachers regarding their role during the coaching of students' knowledge creation processes. Utterance s such as "..You don't really help us coach, we just have to take care of it. In my opinion that is not really being an actor!" (passage 3) versus "Well, the idea was that we didn't want to participate as a teacher" (passage 4) illustrate this tension. Subsequently this resulted in a framing of these conflicting perspectives from the view of the traditional work practices of both groups to create common understanding ".. There is a big culture difference because we are used that everyone is involved You are think as observers" (passage9). Eventually, partners provided suggestions to overcome this tension by a division of labor ".. You could divide one group into two groups […] you can then also ask researchers for feedback during coaching" (passage 11).

The abovementioned developmental tensions set the stage for collaborative analysis and for the creation of a shared understanding to overcome them and change teachers' coaching. Eventually, teachers' coaching transformed towards an increasing emphasis in the collective construction of a planning together with students. This would help students to organize their work and offered teachers a tool that enabled them to monitor students' progress during knowledge construction:

**Phase 1 in coaching knowledge creation**
- Do students have a clear plan?
- What are they eventually going to show, what is their object?
- The teacher has a specific role in this process
- 'Go' or 'no go' decision

*Slide taken from presentation of Teacher3; March 2007*

In this artefact, i.e. presentation provided to other teachers at UniC, teacher3 shows that that the significance of a "planning" is echoed in teachers' coaching practices. This theme can be traced from the tension that team members identified and attempted to overcome earlier. Moreover, teacher3 took up this idea and implemented a 'go-no go' decision in his practice. Then students had to negotiate their planning with their teacher before they were allowed to continue with their knowledge creation projects:

R3 How do you see your role as a teacher now, what is most important?

T3 Well. First that students chose a subject and that they construct a planning. And the task of the teacher is to perform a reality check and argue whether students' planning is a good one or not, to give a 'go' or 'no go' decision at the start. There is where the teacher plays an essential role and this planning gives a good tool for me to observe what is happening and to ensure that students keep in a 'flow' towards the end

*Interview Student group 4; December 2006*

In this excerpt, teacher3 reports that he had adapted his coaching to such an extent that he now asks his students to construct a planning and that it is the teacher's task to decide whether students can continue in pursuing their knowledge construction according to this plan or that they have to construct a more realistic or challenging planning.

Tension resolution of conflicting perspectives about division of labor between members of the design team resulted in creation of a joint venture agreement:

For Utrecht University this agreement involves:
- To perform research at UniC in collaboration with teachers and students concerning the concept of knowledge creation and support thereof
- To realize a long-term relationship between research and educational practice, in which knowledge, insights and experiences are exchanged with the aim of learning and capitalizing from each other

For UniC this agreement involves:
- To obtain more insight and tools to experiment with possible solutions for the challenges and issues which structurally occur in educational practice
- To realize a long-term relationship between research and educational practice, in which knowledge, insights and experiences are exchanged with the aim of learning and capitalizing from each other

*Join venture agreement, first version; April 2007*

## Conclusion

A central challenge in transforming practices of teachers resides in potentials for teachers to learn from and with each other in the work place as they create and advance shared epistemic artefacts relevant for transforming their knowledge practices. To advance our understanding of such processes in teachers' practices, we must explore how these practices evolve. This study focused on developmental tensions between members in university-school partnership, and how their resolution points to practice transformations.

The findings illustrate that interaction between different knowledge trajectories occurred on both the individual and collective platform of the design team and how participants stabilized out of flux by changing their practices accordingly. During meetings practical pedagogical enacted knowledge of teachers intersected with social practices of the educational researchers. At this level, developmental tensions surfaced on the nexus of perspectives, agendas and interpretations of the actors involved in the collaborative design in the university-school partnership. The attempts undertaken to overcome the identified tensions involved the creation of artefacts (e.g. the joint venture agreement) that serve to objectify and afford this transformation.

Tacit knowledge (represented as the network of implicit epistemological beliefs, attitudes and knowledge) was explicated during group meetings and ideas expressed were often taken up by the group and integrated within existent practices, or became the driving force behind the development of relatively new pedagogical practices. For instance, the tension between top–down instruction versus social distancing and differentiation of coaching styles was resolved by a collective envisioning and fleshing out of coaching practices. More specifically, more emphasis was placed on employing students' planning as tools to monitor and to scaffold students' knowledge creation.

In sum, the study shows the transitions observed from identification of tensions, the attempts to overcome them by engaging in transformation of coaching and of social practices towards the creation of artefacts that serve to objectify and afford this transformation. Based on socio-historical perspectives on learning and development, we have appropriated the notion of developmental tensions as a driving force of change and development. Although we do not claim that developmental tensions are the sole impetus of transformations of work practices, the investigation of tensions and their resolution helps to identify the dynamic forces of change and comprise an important constituent and starting point for investigating such processes (Barab et al., 2002; Engeström, 1987; Koschmann, Kuutti, & Hickman, 1998; Murphy & Rodriguez-Manzanares, 2008). For the research reported here, the identification of these tensions provides a starting point for investigating and explaining practice transformations in knowledge creation contexts.

Based on the findings reported in this paper, we have derived a generic pattern of managing or resolving tension, namely: identifying, labeling, framing and solving/suggesting:

1. *Labelling the issue:* Often, tensions are not about the issue at hand (e.g., scheduling a meeting) but rather about what it represents, such as the experience of disrespect or the illegitimate exercise of authority. A tension could only arise as the consequence of one of the professionals in the knowledge construction work to describe a particular problem at hand;
2. *Identifying the tension:* as a result of professionals knowing what the issue is at hand, the contradictions in perspectives, knowledge, attitudes or affects come to the fore explicating the problematic features of the practices under scrutiny. These tensions are explicated in the voices of the several professionals in the collaborative knowledge construction work;
3. *Framing the tension:* subsequently, the tension is framed employing the self-created language, norms and rules of the knowledge community. This framing is necessary for creating a shared understanding of the tension and for constructing a representation of the forces acting in preserving and causing the problematic practices at hand. This will eventually enable professionals to adapt their practices to be able to overcome the tension;
4. *Constructing solutions:* Finally, professionals transform their or shared practices, construct new tools and implement them in the ongoing knowledge construction work.

In future research this generic pattern for tension resolution will be tested in other pedagogical co-design settings.

## Acknowledgements

The present investigation emerged in the context of research and development of Knowledge-Practices Laboratory (www.kp-lab.org, FP6-2004-IST-4, integrated project 27490, 2006-2011) project funded by Information Society Technologies (IST) program of European Community.

## References

Argyris, C., & Schön, D. (1978) *Organizational learning: A theory of action perspective,* Reading, Mass: Addison Wesley.

Barab, S., A., , Barnett, M., Yamagata-Lynch, L., Squire, K., & Keating, T. (2002). Using activity theory to understand the contradictions characterizing a technology-rich introductory astronomy course. *Mind, Culture, and Activity, 9*(2), 76-107.

Dewey, J. (1968). *Democracy and Education: An Introduction to the Philosophy of Education.* New York: Free Press.

Engeström, Y. (1987). *Learning by expanding: an activity-theoretical approach to developmental research.* Helsinki.

Engeström, Y. (1999a). Activity theory and individual and social transformation. In Y. Engeström, M.R. Miettinen & R.L. Punamäki (Eds.). *Perspectives on Activity Theory*. Cambridge: Cambridge University Press.

Engeström, Y. (1999b). Innovative learning in work teams: analyzing cycles of knowledge creation in practice. In Y. Engeström, M.R. Miettinen & R.L. Punamäki (Eds.). *Perspectives on Activity Theory*. Cambridge: Cambridge University Press.

Engeström, Y (2007). From Stabilization Knowledge to Possibility Knowledge in Organizational Learning. *Management learning, 38*(3), 271-276.

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education, 26*(2), 247-273

Hakkarainen, K., Palonen, T., Paavola, S., & Lehtinen, E. (2004). *Communities of networked expertise: Professional and educational perspectives.* Amsterdam: Elsevier.

Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). New York: Harper & Row. (Original work published 1952).

Leont'ev, A.N. (1981). *Problems of the Development of the Mind.* Moscow: Progress Publishers.

Lingard, L., Garwood, S., & Poenaru, D. (2004). Tensions influencing operating room team function: Does institutional context make a difference? *Medical Education, 38*, 691-699.

Little, J. W. (1990). Teachers as colleagues. In A. Lieberman (Ed.), *Schools as collaborative cultures: Creating the future now* (pp. 165-193). Bristol, PA: Falmer Press.

Karlgren, K., Dahlström, A., Lonka, K., & Ponzer, S. (2007). *A new educational annotation tool for supporting medical teams to improve their teamwork and communication.* Paper presented at the ICEM/ILE 2007 – The International Council for Educational Media: Educational Media & Innovative Practices.

Karlgren, K., Dahlström, A., & Ponzer, S. (2008). Design of an Annotation Tool To Support Simulation Training of Medical Teams. In *Times of convergence: Technologies across learning contexts* (Vol. 5192/2008, pp. 179-184): Springer Berlin / Heidelberg.

Matusov, E. (2001). Intersubjectivity as a way of informing teaching design for a community of learners classroom. *Teaching and Teacher Education 17,* 383-402.

Paavola, S., & Hakkarainen, K. (2005). The Knowledge Creation Metaphor – An Emergent Epistemological Approach to Learning. *Science & Education 14*, 535-557.

Reckwitz, A. (2002). Toward a Theory of Social Practices. *European Journal of Social Theory, 5*(2), 243-263.

Schatzki, T. (2002). *The Site of the Social: A Philosophical Exploration of the Constitution of Social Life and Change.* . University Park,: Pennsylvania State University Press.

# Math Engaged Problem Solving in Families

Shelley Goldman, Roy Pea, Kristen P. Blair, Osvaldo Jimenez, Stanford School of Education, Stanford, CA 94395
sgoldman@stanford.edu, kpilner@stanford.edu, roypea@stanford.edu, ojimenez@stanford.edu

Angela Booker, Lee Martin, UC Davis, School of Education, One Shields Avenue, Davis, CA 95616
anbooker@ucdavis.edu, leemartin@ucdavis.edu

Indigo Esmonde, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, M5S 1V6,
indigo.esmonde@utoronto.ca

**Abstract:** Research indicates that people engage in rich mathematical practices in everyday activities, yet little is known about school-aged children's mathematics learning within the family context. This paper reports results of an interview study with 20 families to understand contexts and activities that engage mathematics in the family setting. The results indicate that problem solving is frequent activity, and that mathematics is engaged in accomplishing problem solutions in a range of contexts or situations. We describe features of math engaged problem solving and describe how it is value driven. We see multiple kinds of math and multiple people drawn into problem solving, and we identify socially distributed mathematical practices. These findings implicate the family as an under-recognized, yet rich source of math teaching and learning.

## Introduction

We investigate the diverse contexts and activities in which middle school age learners and their families engage in mathematical practices and problem solving. We describe several features of problem solving for the families we studied, characterizing the structure of their mathematical activities, and analyzing the social conditions and arrangements for their family-based mathematical practices. Several important features of problem solving around the home emerge: (1) the family is a setting giving rise to problems to be solved, which also generate mathematical activity; (2) values steer problem definition, identification, solutions and their judged adequacy; (3) different types of math and math at different levels come into play concurrently; and (4) it is critical to view mathematical thinking and learning as embedded in activity and in a social, historical, and distributed knowledge environment.

We first give some background on our approach and describe our research methods. Then we discuss each aspect of familial problem solving and present one example for each that is characteristic of the feature. We follow the descriptions with a summary and discussion of the significance of this characterization of problem solving with mathematics.

## Background and Perspectives

Understanding mathematical problem solving in the complexity of family life activities is supported by certain orientations to theoretical and analytical approaches. Work on problem solving has had a long history, with significant roots in Gestalt psychological studies by Wertheimer (1945), Duncker (1945) and others, which had direct influences on the cognitive revolution in studying thinking (Newell, 1981; Simon, 1999). The mathematician George Polya in *How to Solve It* was another major contributor to conceptualizing problem solving processes in mathematics (1957). None of these approaches, however, sought to understand the situated nature of mathematical practices in relation to everyday life activities. In the 1970's, spurred by everyday cognition theoretical formulations from Neisser (1976), Bronfenbrenner (1979), and Cole (1971), an ecological orientation to looking at mathematics use *in situ* was developed. It was advanced throughout the 1980's by Lave (1988), Saxe (1988), Nunes et al. (1989), Scribner (1984), and others more recently (Goldman & Booker, 2009; Hoyles et al. 2001; Nasir, 2000; Satwicz & Stevens, 2008; Stevens et al., 2006). We know from such studies that people engage mathematically in the course of everyday activities from dieting to shopping to sports to the work of nursing and dairy order filling.

We take a view of problem solving that recognizes its complexity and finds problems being generated in many life situations. Problems and problem solving are part of the social fabric in peoples' lives, and they are distributed in time, place and across persons. They involve cognitive, social, cultural and physical means in their definition, generation and solutions, and a direct pursuit of both outside constraints and peoples' desires. We are especially interested in the problem-solving that people do during their family-based activities, how it both requires and generates mathematical practices, and how these mathematical problem-solving practices instantiate learning.

We see mathematics practice as ubiquitous in people's family-based activities. We also take a comprehensive view of learning that is based on activity in practice (Lave & Wenger 1991), and consider the family as a little examined yet significant cultural environment for activities and social relations that constitute a sizeable proportion of human experience and involve poorly understood forms of learning (Bronfenbrenner, 1979; Leichter, 1975). In relation to math, Ginsburg and his colleagues document how and what mathematics young children are "taught" at home (Ginsberg et al., 1999). Jackson and Remillard (2005) and Rogoff, Ellis, and Gardner (1984) show that the explicit teaching softens and changes once children enter school. The literature leaves us with a very obscured view of the significant role that families might play in mathematics teaching and learning, yet with the hopes that math engagements might be ubiquitous in familial circumstances and activity. Our work enters the family space to better understand the significance of the family as a setting for mathematics engagement, learning and the development of quantitative literacies (Steen, 2004).

## Methods

The Family Math research project investigates the ways families with middle school aged children characterize their engagement in mathematical problem solving during the course of individual and family activities and events.[1] Our goal is to identify the social, cultural, and material contexts that are relevant to and create opportunities for mathematics learning and practice as well as to characterize similarities and differences in family math situations. We seek to identify the resources family members use for solving problems together, characterize the structure of these activities, describe the resources used in solving math-related problems, and analyze the social conditions and arrangements for family-based mathematics practices. Eventually, these understandings of mathematics of family life will help the field identify and design for points of synergy for strengthening the role of the family in mathematical teaching and learning and to build bridges between math in the home and math and school.

In the results reported here we present data based on two-hour interviews with 20 families providing narrative accounts of math in their daily lives. Narratives provide us with participants' accounts in their own words and allow for us to have records of how people relate stories about their math related activities, including family members self-identifying as African-American (N=5), Latino (N=23), Asian (N=7), Pacific Islander (N=8), Multi Racial (N=6) and white (N=25). The families also represent a spectrum of economic diversity, ranging from low income (children can receive free lunch) to upper-middle income families. The educational levels of the heads of families was also diverse, ranging from high school education through graduate school.

### Data collection

We developed a semi-structured interview to have people discuss the activities they engage in as part of family life, work and school, allowing them to give us particularized versions of how they think about and accomplish their life tasks. The interview gave family members ample space for describing their experiences, while also collecting certain information from all families such as education levels, perceived comfort with mathematics, school attainment, number and description of math classes that parents (and students) have taken, and their attitudes towards mathematics in general and towards school math. The protocol included two segments to provide parity of data across all families. Each family member told a "Math in a Minute" story about an experience of any kind, positive or negative, alone or with others, that they had with mathematics. Each family also was asked to choose a cell phone plan by using brochures from four providers we provided that would meet the needs of the family. This task put the family into action and provided observable, interactive accounts of the family in a problem-solving situation. All interviews were videotaped and took approximately two hours.

### Data Analyses

The analyses of the interviews proceeded through several stages and treatments. Three were particularly critical to the analyses of data for this paper: content logging, identification of talk and interactions involving problem solving, and development of a Math Inventory. These three data examinations eventually became the basis of a database on the intersection of problem solving and mathematics. All videotapes were first content logged to delineate events, turns in conversation topics, stories, events, and sub-topics of interest inside participants' accounts. Analyses of the segments resulted in 14 general categories, such as formalisms, gestures, props, efficiency, (math) avoidance, emotions, division of labor, and values. This paper emphasizes the categories of "problem-solving" and "type of mathematics" so we could better understand the occasions where family members depicted a situation *as creating a problem for them* and for which they engaged mathematics. These categorization processes resulted in a database comprising 315 stories—events that participants described where they were solving a problem using one or more types of mathematical practices.

We also developed a list of 20 categories for "types of mathematics used," establishing definitions with exemplars from our dataset for each. To consider overall patterns in our dataset of problems involving math, we examined the frequency with which the various types of math occurred within the various family activities.

## Findings: Four Features of Math Engaged Problem Solving in the Family

Four aspects of math engaged problem solving are described through vignettes based on our families' accounts of their math engagements. We tell their stories and point out the ways in which they are representative and characteristic of patterns we saw in the larger data set.

### I. Everyday problems lead the math

The most overarching feature of family problem solving is that problems come first, with math following. The complexity of family life and many interactions among family members and other institutions generate many contexts that require problem solving. The problems are complex (a need to renovate a bathroom), involve many steps over significant time periods (training for a race), and often are nested inside larger or connected problems (playing an extended game). Problems are defined and constrained by life imperatives (we must pay our credit card bills).

The knowledge and solutions derived from problem solving must align with situations and real constraints. People generally define their problems, and secondarily consider how to solve them, and then when to use math. Math rarely leads (except when family members directly want to learn or practice math). Single problems expand to fill the space, time and solution resources available. They may branch, with the resolution of one problem spawning new or next problems. People talked of the nesting and cascading of problems to be solved. We saw great examples of this in home improvement projects. People seem to meet up with problems once they start improvements, only to realize they must be solved in coordinated or timed ways. This is not to suggest that all problems are like runaway trains or that every problem facet being solved engages mathematics. It is to suggest that problems to be solved are complex ones that often require shifting attention, coordination of multiple nested activities, and unplanned-for triage processes.

When is mathematics brought to bear on problems in the rough and tumble of problem solving? Sometimes people are aware they are bringing math to the situation; other times they are engaging unawares of their mathematical practice. We had many instances in the database where we coded that the math was inferred—that people could not have accomplished what they told us they did without doing some mathematics. In this way, math could be *sub rosa*. The problems to be solved are often unique and situation dependent, people are generally required to evaluate their own solutions, and if correct, decide if the correct solution is really relevant or appropriate to the particularities of the situation. For example, people described both "doing the math" and "tossing it" when the idealized math did not fit the local situational constraints. We can see some of these characteristics of problems in a family's condo renovation project.

In one family, mom Yali and daughters Vivian and Renee are in the process of renovating a condo that Yali recently bought. Vivian is in law school and Yali recently became a teacher after a mid-career change. They want to save money on the renovations and Vivian likes carpentry, so she is managing the renovations and doing some of the work. Vivian describes how there are so many decisions to make—moldings, the placement and the size of walls, cabinets, which way to place the tub, where to put the sink? Closets, should there be one or two? Vivian discusses how they conceived of the renovation.

> *Interviewer*: Did you use any tools or ways to visualize to make these decisions? How did you actually decide where to put things?
>
> *Vivian*: It was a combination of just paper, like drawing a two dimensional…this is a square that is the kitchen. Where are the things going to be? And then actually showing up on site. I had to go down there a number of times, which was the hardest part, because, you know, we're trying to figure out, okay, we have this space between the tub and the toilet and how much space do you need for your knees? So we would just take the tape measure, and we're like okay, we want the [height of the] bench by the dining table. And we're like, well, how high is the bench by the dining table? There are these various measurements. So we literally took the tape measure and measured our knees from the floor sitting on [the bench]. So then you want to design, okay, what are our options? And you draw those on paper and then you actually show up and then you sit on the toilet. How many inches do I need, if the tub comes to here will I be able to sit down? <Yali laughs> We will, but if we sell the house to someone who's taller, will they? So it was kind of a back and forth.
>
> *Interviewer*: So you used your body for some of that?
>
> *Yali and Vivian*: Oh Yeah…

> *Interviewer*: …And you drew plans. When you drew the plans did you draw them to scale? Or did you just free form draw them?
>
> *Vivian*: I did both. Some things, well I'd mostly draw the free form just to give to the contractor an idea, okay this is what I decided. The tub is going to go here and I'd draw the inches...But I actually got on my computer and found some grid to work with to actually get more of the proportions. I didn't have a ruler in my whole room. So I couldn't even just measure a piece of paper and draw inches myself. I had to do it all online.

The family enjoys that they measured with their bodies, and upon Vivian's telling of the story, they all move around and giggle. Vivian's story on managing renovations reveals how informed critical decision-making is to their home improvement project. Vivian uses math in this work. She draws plans free form, but also draws them to scale in order to communicate with the contractor. The math involves knowing the dimensions of the large fixtures and pieces and the dimensions of the spacing they are being placed in. Drawing the scale version of the plans involves assigning scale units and engaging in proportional reasoning and problem solving. If a bathtub is 60 X 30 inches, how big is that on this particular graph paper? Still, even scaled plans might not be enough, and the women use their bodies to give a reality check to the calculations to make sure what is accurate on the 2D scaled paper plan will also work in the 3D world of real people moving around in the spaces. In the renovation project, daughter Vivian knows not only how to make and use scale models, but also what their limitations are, and when to switch into reality mode. This back and forth cycling of model to reality is interesting, and typical in this type of situation.

The renovation also included a new kitchen and new closets in the entranceway. There was a lot of figuring and fitting going on. To measure for the stone countertop, Vivian entices thirteen year old Renee into realizing she could use the Pythagorean theorem to complete the measurements for the counter in the stovetop corner. There was also clever financial planning around emerging problems as they were discovered, from having to convert from metric to standard measurements to figure out if the cabinets would fit before ordering, to balancing loans for the construction costs against 0% financing on appliances and fixtures, with the projection of a well-paying job when law school was over in a year. The problem space was huge, cascading and needing attention over an extended period of time. The math identified in the renovation story was extensive as well and included arithmetic, comparing magnitudes, geometry, measurement, optimization, unit conversion; proportional reasoning, decimals, unit conversion, formula, measurement, estimation, data representation, and interest rates.

## II. Values steer the problem definition

*Why* engage all of this problem solving? Our analyses revealed that exploring values answered that question. Why bother to take a student loan to help finance your mom's condo renovation and juggle a huge amount of debt for several years? Values also answer the *how* question. Vivian could hire someone to design the new kitchen and bathroom, or she could figure it herself, especially since she had lots of ideas and she thought she knew enough to get started. Values drive some of the decision-making around problem solving, and in volatile emotional moments, steer people into certain practices and solutions (instead of arguing, let's check the figures). Peoples' preferences and their values actually are determinant in whether or not a problem will even be engaged (Pea & Martin, 2010).

When families spoke about their problem solving strategies and gave examples of the kinds of problems that arose in their lives, family members consistently talked about what they valued. Families often had to consider the factors that mattered when making decisions. Was getting the "best" deal most important? Did social relationships play a part? Was time an important consideration? Was it important to make decisions by taking into consideration *all* of the relevant information? Each family's answers to these questions varied, sometimes even from one problem to the next. The answers to these questions often organized the problems to be solved by defining the constraints, the resources, and the criteria by which the outcomes would be judged. Values helped families determine *how* to solve problems once they were identified as such. When one mother wanted a savings plan to pay down credit card debt, she chose a socially regulated solution that traded the "best" financial deal for the power of social commitment and cultural participation.

One family's approach to purchases illuminates not only how values enter into the mathematics of their decision-making, but in this case, the creation of a new kind of value exchange unit (product goods vs. days mother has to work away from home). The dad, Jim, is an analyst and the mom, Liz is a part-time financial consultant. Liz works part-time because she tries to stay at home with the children as much as she can. She does take consulting jobs to keep up with expenses, but she tries to optimize her time at home with what she can earn through consulting gigs. When we asked how they handle the purchase of big-ticket items, they responded by going straight to the importance of their values. Jim said that he and Liz had been thinking about "bigger ticket item" purchases:

> We have got into an interesting concept about what the value of bigger ticket items is to us personally, and value is a very personal thing. Since my wife is a consultant and works on billable hours, we have gotten into this thing of looking at something, and if it is a big ticket item, thinking about how many hours it costs for us to get it. Whatever it is: I have to work three days, or it will cost me three weeks of effort if I want to do this, whatever the big ticket item is. And we sit back and we say, you know: is this really worth three weeks? No. Is this worth three days? No. So it depends – and sometimes – is this worth three days worth of effort? Yes. It's an interesting way of thinking of things, and trying to assess the value, which we do on bigger ticket items… We oftentimes say, you know, it's not worth that to me. I am not going to work five days to get this thing. It's just not worth it to me. It is an interesting value concept that just the two of us think about.

The family actually pits Liz's time at home against their wants, knowing they can have one or the other. Liz's home time usually wins, and the family goes without. This is especially so if the purchase was a flat screen TV. However other kinds of big ticket items, such as paying to offset the tuition for their middle school aged daughter's private school, is considered worth the consulting hours. Jim and Liz went on to tell us that even their children have come to use a similar approach in their own thinking. It started as the children were getting to the ages where they wanted to purchase souvenirs when the family went on a vacation. They decided to give their children $10.00 and let their children budget their own purchases, forcing them to ask the question of whether what they would purchase was worth it.

> Jim: But it has translated down a bit with the kids and the money that we give them on vacation, and we…they have ten dollars, and then they go through this value exercise to say 'do I really want that candy bar for two dollars or not?' And they are starting to make value judgments themselves on what is important to them, and whether or not they are going to spend the money on a particular item, or save it.

What such a structuring of the values undergirding decision making for bigger ticket purchase items exemplifies is that how a family weighs, or creates equivalences between, different categories of resource (incremental earnings, big ticket purchases, family time), affects the decisions that they make in how resources are allocated. In turn, features of the ideology, as well as of the mathematical practices, can serve as a model that the children emulate in their own age-relevant versions of the adult problems. For example, the math involved in Jim and Liz's purchase to consulting hours trade off includes unit conversion, arithmetic, and proportional reasoning. As the parents discuss family purchases, the children are exposed to mathematical reasoning involving these concepts, as well as the value considerations their parents consider important.

## III. Different maths engaged in service of problem solving

Two findings relating to the math family members engaged while problem solving are of special interest. The first is that a number of problems in our database indicate that people engaged multiple maths on the way to problem resolution, even when we exclude simple arithmetic or computation from the data set. The second is that, even for the same problem, family members approach problems with different ways of bringing math to the solutions. There are several conditions for multiple maths to co-occur. People in the family sometimes approached the same problem in different ways. In some of the cases, they knew the way of another family member, but chose to use their own. In a few cases, this led to circulation of a problem from one person to the other, and in these instances, we found family member playing to perceived expertise of others.

Engaging multiple kinds of math to problem solving is a very interesting and positive aspect of family mathematical practice, and we see it in stark contrast to how problem solving is taught in school math. In school, the approach to math most preferred is a development-sequential one. Math in school tends to concentrate on learning and practicing one kind of math at a time, and this is generally the case, even when problems are of an applied nature. In school, the applied problem is often designed to practice a singular math concept; at home, the problem is generally designed to accomplish a goal that is desired and sought, often with the engagement of a variety of math practices for accomplishment. The family offers children the opportunity to develop mathematical practice under conditions that matter to them.

For one middle school boy, Ravi, a pre-algebra student, opportunities to apply his new problem-solving skills abound at home. His mom talked about how he constantly enjoyed trying to calculate the best deal. The family told the story of Ravi's part in helping them decide what car to purchase. The family was looking at various Toyota models, and they were considering a number of factors in making the decision. Ravi compared the options and saw gas mileage as the deciding factor:

> *Ravi*: Yeah, uh, so we were considering between a Toyota Corolla, a Toyota Camry, and a Prius. And I said that we should buy a Prius because my mom drives a lot of miles [Mom &

Son laugh]. And so, uh, with a Prius, uh, you get, uh, I believe you get some tax benefits, car, auto insurance [Mom & Sister smile].

*Dad*: [returns, sits down and pats son on the back while smiling and laughing] You don't need to know that.

*Ravi*: … and, a Prius, um, is better for the environment also.

While Mom and Dad would also take the tax benefits into account, they knew Ravi was not yet ready to weigh those benefits. Still, he was intent on making a case for the mileage in terms of environmental effects along with cost savings at the pump and at income tax time. He had figured the differences and savings for his parents and argued for the ways the pump and environmental impact savings would even out the extra purchase price. His interest in optimizing was balanced by his knowledge of his mom's commute and a general appreciation in the family for thoughtful, mathematical arguments. Ravi demonstrated how he put what he learned in school to work for him on the Prius problem:

*Ravi*: Yeah, like um, usually in the car, like uh, for the Prius, like miles per gallon. When you see 50 miles per gallon, that means if you drive, you will use 1 tank of gas if you drive 500 miles. That means, if you fill up your tank you can drive from here to LA in that tank.

In Ravi's example, he set up a problem and argumentation for his choice that would put the mileage benefits of the Prius into perspective. He knew he could make it from Palo Alto to LA on a 10-gallon tank. It is important to recognize what drives Ravi to set up a proportional calculation. He wants to find the best of three car options for his family's needs. His interest in optimizing gas mileage to get his family the best financial deal while making a responsible environmental decision works nicely with his penchant for math problem solving generally. The car purchase context sets up the opportunity for Ravi to draw on several types of math problem solving at once. The car purchase demonstrates the way Ravi is able to draw on his school knowledge inside a relevant family context. He tackles everything from arithmetic and measurement to optimization and proportional reasoning in order to support his argument.

## IV. Mathematical thinking as activity in distributed knowledge environments

The stories that families told us indicated that math engaged problem solving was most often a social affair. This was so even when we asked an individual for a story. Children often mentioned how siblings, parents, grandparents, or friends were part of their math stories. Adult problem solving often involved others, even if they did not always involve children. Social, historical and cultural practices were deeply embedded in problem solving. Playing math games with a grandfather, working on puzzles as a family, inventing fantasy businesses for play, and saving money and managing debt were all deeply social and distributed across persons, time and setting. When math was involved, it became entwined with the persons in activity. Math engaged events, like other learning situations, were embedded and distributed among people, artifacts, activities and settings (Gonzales et al., 2001; Pea, 1993). Individual accounts of how the math practice was distributed differed, yet we found only few instances in family-based activities where math engagements were isolated or individual.

Mathematics during a road trip plays out in one family for whom mental math is key, particularly for the father, Steve. Throughout our interview, Steve's use of mental math to figure out problems is a common theme recognized by the rest of the family and becomes a source of amusement. For example, Steve describes doing "nutrition math" around each meal to stay on the *Zone* diet and also decrease the amount of cholesterol he and his wife consume. Such activity involves calculating protein grams, different forms of carbohydrates, and fat per meal, as well as the amount of cholesterol and total number of calories. According to Steve, "you almost need a computer to do the nutrition analysis of each meal at this point. So, I have this much saturated fat, so if I use olive oil instead of this oil…a lot of [calculations like] that." However, rather than using a computer to keep track of the calculations, Steve says he adds up the different quantities in his head as he builds a menu. His comment - "I do it in my head" - causes laughter from his wife, Sharron, and daughter, Elise. Sharron jokingly asks us: "Do you see a theme here?" - referring to the father's focus throughout the interview on doing calculations in his head. Sharron and Elise also figure cholesterol, but with different methods—one cooks with no fat, since "zero times zero is zero", and the other computes cholesterol with paper calculations.

The focus on mental math is not only important to Steve, but it is also passed down to Elise, with the involvement of the whole family. Elise describes a game she played in the car on a trip with her parents:

*Elise*: We were coming home across the Golden Gate Bridge, and I wanted to see if I could hold my breath across it. Our guidebook said it was 2 miles long, and at 45 miles/hour, so we had to figure out how long I had to hold my breath, and it ended up being 3 minutes. And that's a long time, so… I didn't hold it for the whole entire time.

*Sharron* (chimes in laughing): We drove faster!

> *Steve* adds: We realized we'd have to change some of the physics in order for this to work, since 3 minutes was too long. So we drove faster and shortened the span to just the part that was kind of out there. She figured out the stuff while we were driving. That was her project to do the math.

This game was initiated by Elise, as holding her breath across a bridge was something she had done on a fieldtrip with her friends. However, it was taken up by the whole family, and became a problem to solve by the family unit. According to Sharron, for Elise to figure out whether she would be able to hold her breath "was a good exercise that took part of the time between Santa Rosa and the Golden Gate Bridge." Once she figured out that she would not be able to hold her breath long enough, the family unit took it on to figure out how to help her succeed, both by changing the constraints of the problem, the factors involved, and what would count as a solution. As a family, they decided that she should not try to hold her breath for the entire bridge, but rather concentrate on a shorter span of the bridge that was more manageable. The father also changed the speed of the car, speeding across a section of the bridge to make it easier for his daughter to succeed. This episode was recounted by the family with enthusiasm, with all members actively chiming in to the description. What started out as a game for the daughter became a challenge for the family to solve.

This family focus on mental math is not new for Elise, and she gets practice in many situations with her parents' help: "There are a lot of times in cars when dad has a problemish thing, like if we're going this fast, how long will it take us to get somewhere." Steve adds, "That's right. And you have to do it in your head. Which is hard, isn't it? But really good practice." At the store, Elise's parents often ask her to help with the "shopping math," asking her to do price comparisons to figure out which is the best value, as well as figuring out the nutritional analysis of foods based on the serving size they actually eat. In talking about shopping math, Steve says:

> They try to teach rounding in the schools obviously, but when [Elise] gets a problem that we would just instinctively round off, because we think, 'I'm never going to calculate 16.3 divided by 8.7' or something, she's instinctively trying, you can see, especially - that's why it's good to try to do stuff in your head – try[ing] in her head to do 16.3 divided by 8.7 (Steve gestures as though writing out a long division problem.) And it's like, 'ok, this is a good time for rounding'. And she's like 'Oh, well then, 16 divided by 8. Now it's easy'. But it's not an instant conclusion that you leap to.

Sharron adds: "probably more than anything, that ability to get common number sense for everyday use for quick, (snaps) quick answers just seems to be really key for her, and she's getting there." Her family is explicit about using rounding as a strategy, as Elise recognizes. She says about Steve: "You do that a lot. I have trouble with a math problem. And then he says, well what if the numbers are this and this, could you figure it out then? Then, it's oh, that's how you do it."

## Summary

Four features of mathematically engaged problem solving were found across our interview data: Mathematics was engaged as people solved problems in their lives; problems that received attention did so because people valued the end goals related to the problem solving; multiple maths were brought to bear on problems; and, math was embedded in the socially distributed learning opportunities. Mathematics followed as people attended to other goals. These features varied across families, situations, and events, to the extent that family position, histories, styles and cultural practices had a lot to do with how and when problems and math were engaged.

These results have several implications. First, rich problem solving in the family is not appreciated, and family mathematics even less so. If mathematical engagements are recognized, they are usually considered to be of less consequence than the more formal math structures taught in schools. Second, the results of our analysis indicate that family mathematics practices are socially embedded in some of the families' most important and valued activities. Family members engage with math because they need to or want to. They support each other through reliance on, and knowledge of, their distributed expertise. Third, children are often witnesses and participants in these vital family problem-solving contexts, and the mathematics engaged is definitely providing learning opportunities—about math and life. Fourth, the mathematics covered has alignment with school math objectives through pre-algebra—another reason why the learning opportunities available in families are rich ones.

## Endnotes

(1) The Family Math Study is conducted as part of the LIFE Center (NSF REC 0354453). Eventually, its results will also inform a range of studies about learning in both formal and informal learning environments (for more about the LIFE Center, http://www.life-slc.org).

## References

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Cole, M., Gay, J., Glick, J., & Sharp, D. (1971). *The cultural context of learning and thinking: An exploration in experimental anthropology.* NY: Basic Books.

Duncker, Karl. (1945). On problem solving. *Psychological Monographs,* 58, no. 270, i-112.

Ginsburg, H. P, Inoue, N., & Seo, K. (1999). Young children doing mathematics: Observations of everyday mathematics. In J. Copley (Ed.), *Mathematics in the early years* (pp. 88 – 100). Washington, DC: National Association for the Education of Young Children.

Goldman, S. & Booker, A.N. (2009). Making math a definition of the situation: Families as sites for mathematical practices. *Anthropology & Education Quarterly*, 40:3, 369-387.

González, N., Andrade, R., Civil, M., & Moll, L. (2001). Bridging funds of distributed knowledge: Creating zones of practices in mathematics. *Journal of Education For Students Placed at Risk*, *6*(1&2), 115–132.

Hoyles, C., Noss, R. & Pozzi, S. (2001). Mathematising in practice. In C. Hoyles, C. Morgan & G. Woodhouse (Eds). *Rethinking the mathematics curriculum.* London: Falmer Press. 48-62.

Jackson, K., & Remillard, J. T. (2005) Rethinking parent involvement: African American mothers construct their roles in the mathematics education of their children. *School and Community Journal*, 15(1), 51-74.

Lave, J. & Wenger, E. (1991). *Situated learning*. Cambridge: Cambridge University Press.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life.* Cambridge: Cambridge University Press.

Leichter, H. (1975). *The family as educator*. New York: teachers College Press.

Nasir, N. (2000). Points ain't everything: Emergent goals and average and percent understandings in the play of basketball among African-American students. *Anthropology and Education Quarterly*, 31(3): 283-305.

Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: WH Freeman and Company.

Newell, A. (1981). Duncker on thinking; An inquiry into progress on cognition. In S. Koch & D. Leary (Eds.), *A century of psychology as science: retrospections and assessments.* NY: McGraw-Hill.

Nunes, T., Analúcia D. Schliemann, A.D. & Carraher, D.W. (1989). *Street mathematics and school mathematics*. Reston, VA.

Pea, R., & Martin, L. (2010). Values that occasion and guide mathematics in the family. In W. R. Penuel & K. O'Connor (Eds.), *Learning research as a human science: National Society for the Study of Education Yearbook*, 109(1).

Pea, R. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognition* Cambridge, England: Cambridge University Press. 47–87.

Polya, G. (1957). *How to solve it: a new aspect of mathematical method* (2nd ed.). New York: Doubleday.

Rogoff, B., Ellis, S., & Gardner, W. (1984). Adjustment of adult-child instruction according to child's age and task. *Developmental Psychology, 20*(2), 193-199.

Satwicz, T. & Stevens, R. (2008). Playing with representations: how do kids make use of quantitative representations in video games? *International Journal of Computer Mathematics Learning, 13*, 179–206.

Saxe, G., (1988*).* Candy selling and math learning. *Educational Researcher, 17* (6).14-21.

Scribner, S. (1984). Studying working intelligence. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 9-40). Cambridge, MA: Harvard University Press.

Simon, H.A. (1999). Karl Duncker and cognitive science. In *From past to future, 1(2). The drama of Karl Duncker*, pp. 1-11. Worcester, MA: Clark University Press.

Steen, L. A. (2004). *Mathematics and democracy: the case for quantitative literacy*. Washington, DC: National Council on Education and the Disciplines.

Stevens, R., Mertl, V., Levias, S., & McCarthy, L. (2006). Money matters: The social and material organization of consequential financial practices in families. In S. Barab, K. E., Hay, & D. T. Hickey (Eds.), *Proceedings of the International Conference of the Learning Sciences 2006: Vol. 2*. (pp. 1088–1093). Mahwah, NJ: Erlbaum.

Wertheimer, M. (1945). *Productive thinking (1st edition).* New York: Harper & Row.

## Acknowledgments

We'd like to thank all of the families that participated. They enlightened us about both problem solving and math. Ana Rios and Judit Moschkovich of the NSF CEMELA Center conducted a few of the interviews, as did Ernesto Colin. The findings and opinions are those of the authors and do not represent the opinions of the NSF.

# Transformative professional development: Cultivating concern with others' thinking as the root of teacher identity

Rachel E. Scherr and Hunter G. Close, Seattle Pacific University, Seattle, Washington, 98119,
rescherr@gmail.com, hclose@spu.edu

**Abstract:** At the university level, science education reform is limited by novice instructors who do not share reformers' educational values. The current challenge is to provide professional development (PD) that will be *transformative:* rather than merely exposing participants to reform instructional materials, it will change what future instructors believe is worthwhile in science education. Our view is that professional development is most effective when it enriches the intellectual life of teachers and builds teacher identity by cultivating concern with student thinking. In an example of an activity that originates in this perspective, participants interview peers about physics concepts and conduct structured reflections on the interviewing experience. The activity holds promise for improving participants' abilities to listen and to teach effectively, increasing their esteem for the intelligence that teaching requires, and increasing their interest in K-12 physical science teaching as a profession.

## Major issues addressed

Physics education research has produced a wide variety of high-quality instructional materials for use in introductory university-level courses (Laws, 1999; McDermott & Redish, 1999; Sokoloff & Thornton, 2001; Thornton & Sokoloff, 1997), including tutorials (McDermott, Shaffer, & PEG, 1998). The limiting reagent for physics education reform at the university level is now implementation (Tobin & McRobbie, 1997).

### At the university level, reform is limited by novice instructors who do not "buy in"

Experienced educators and developers are well aware that successful implementation of reform instructional materials includes establishing appropriate norms for learning in the classroom (Henderson, Yerushalmi, Kuo, Heller, & Heller, 2007). At the university level, for example, science education reform has often taken the form of collaborative worksheets ("tutorials") used in recitation sessions. In these classrooms, norms include an emphasis on conceptual understanding (and a concurrent de-emphasis of algorithmic application of formulas); an expectation that this understanding is best achieved through explaining one's own thinking, listening and responding to others' ideas, and constructing arguments; and an acceptance of instructors as facilitators of this process rather than sources of correct answers. The establishment of these norms is "among the most critical and subtle features of implementing these reforms" (Finkelstein & Pollock, 2005).

From the students' point of view, the graduate teaching assistants (TAs) and/or undergraduate learning assistants (LAs) who lead each tutorial section are important arbiters of these norms and expectations. The development of these norms by these novice instructors is thus a critical task of tutorial implementation Instructors who "buy into" tutorials are more likely to convey their respect for the material and the tutorial process to the students, as well as learning more themselves. This development is nontrivial: although graduate and advanced undergraduate students may be presumed to be more sophisticated learners than introductory students, they are in some cases more thoroughly embedded in traditional teaching practices. Prior research by the authors on tutorial implementation has provided examples of how lack of buy-in by an individual TA can undermine the effectiveness of tutorials (Goertzen, Scherr, & Elby, 2009).

Lack of buy-in can have roots in impoverished beliefs about the intellectual life of teachers In our culture, the work of a science teacher is not regarded as that which requires the great power of a creative and analytical intellect. Novices who believe that an instructor's job is merely to "check that students get it" may be frustrated with tutorials, in which instructors are expected to facilitate student exploration of a variety of solution paths. The public perception of science teaching as minimally intellectual work diminishes the interest the next generation of talented thinkers might have in doing the work: talented, intelligent students want to grow up to do something that they believe requires talent and intelligence, and teaching is not usually considered to be one of those things. Novice instructors' lack of appreciation for the intellectual rigor and creative opportunities in teaching impedes science education reform.

### Typical professional development rarely helps instructors value reform instruction

TAs and LAs receive various forms of professional development (PD) associated with their teaching responsibilities. While these programs may provide novice instructors with necessary information and general skills, few help them recognize the worth of reform instruction. Some novice instructors participate in workshops and seminars focused on classroom management, grading, facilitating discussion or learning questioning skills (Gilreath & Slater, 1994; Hollar, Carlson, & Spencer, 2000). These types of PD are oriented toward providing instructors with specific skills and information, and are not specific or sustained enough to have much effect on their values. In a tutorial program, the most common form of professional development is a weekly session aimed specifically at preparing instructors for the coming week's instruction. The primary activity is usually to complete the tutorial worksheet as if they were students. This specific preparation is surely necessary to effective instruction. However, research indicates that it is not sufficient to help the participants appreciate the power of tutorial teaching. In particular, the social and environmental context of the tutorials – including classroom, departmental, and institutional levels of implementation and support – has been shown to strongly affect whether TAs buy into tutorials, and probably outweighs the influence of the typical preparation activities that we prepare for them (Goertzen et al., 2009).

Some novice instructors take courses that offer instruction in pedagogical content knowledge and constructivist learning theories (Hammrich, 2001; Ishikawa, Potter, & Davis, 2001; Lawrenz, 1992; McGivney-Burelle, DeFranco, Vinsonhaler, & Santucci, 2001; Otero, Finkelstein, McCray, & Pollock, 2006) Where pedagogy courses are offered, there is more potential to help instructors value reform teaching, because the courses address models of teaching and learning directly for a sustained period of time. A pedagogy course is particularly likely to be effective when there is good alignment between the learning theories presented in the course and the practical teaching experiences that the instructors have: if the instructors learn constructivism in the pedagogy course, but are permitted or requested to teach by telling in the classrooms that they lead, they are less likely to appreciate constructivist teaching methods. Assessment of professional development that includes a pedagogy course should thus include systematic observations of the participants' classrooms, not only surveys or interviews.

## Potential significance of the work

The current challenge for university physics education reform is to provide professional development (PD) that helps instructors to appropriately value reform instruction. Such PD will be *transformative:* rather than merely exposing participants to reform instructional materials, it will change what future instructors believe is worthwhile in science education. We take the perspective that this transformation is a *philosophical* and *epistemic* one, involving a change in what is thought to be worth knowing.

### Transformative professional development enriches the intellectual life of teachers

In the transition from Plato's philosophy of the natural world to Aristotle's, phenomena of the natural world were promoted in epistemic status. In Plato's philosophy, the actual world is not something that is worthy of study in and of itself; it consists instead of a variety of imperfect copies or images of ideal principles These ideals are the "real" things of concern, and the proper subject of contemplation and analysis, while their shadows and distortions (of which the world we experience consists) are not. In contrast, Aristotle's philosophy (which followed shortly thereafter) granted a new and higher status to the world of phenomena, as we experience it. This world is populated with all sorts of things that are themselves worthy of study, as they are. On this, Aristotle says (Aristotle, c. 350 B.C.):

> *It is the business of experience to give the principles that belong to each subject I mean for example that astronomical experience supplies the principles of astronomical science; for once the phenomena were adequately apprehended, the demonstrations of astronomy were discovered. Similarly with any other art or science.*

An Aristotelian might see the Platonist who does not care to pursue truth by first studying the instantiated world as fixated on the world as it should be, rather than the world that is.

A similar dichotomy of perspective exists for novice instructors learning to think about student thinking. Most of these instructors are very (and rightly) concerned with understanding physics as it *should be* understood. This is not to suggest that there is only one way to think correctly about physics; surely, there is not But novices are usually very aware that they have ways of thinking that they may want to avoid, and ways of thinking that they strive to use more consistently and reliably. They are right to recognize that they have not yet mastered basic physics (though they probably also underestimate how long it takes most physics instructors to do so). Thus, initially, novice instructors are generally *fixated* on correct understanding in physics – on their own part, on the part of their peer instructors, and on the part of the students that they would teach. This is not to say that their attention is so focused on correct understanding of physics that they always understand clearly for some bit of instruction what the target understanding is; it means that correct understanding is their primary *concern* during teaching. They are fixated to the point of failing to understand expressions of student thinking in physics as phenomena worthy of study and explanation, as they are, and not only secondarily, as they fail to match up with correct understanding in physics. In this way, beginning instructors are Platonists when it comes to alternative ways students have of thinking and knowing. The most important thing to observe (according to the beginning instructor) about student thinking when teaching is how that thinking is wrong, and since this is teaching, an activity in which the student (we assume) is supposed to be changed by the interaction, the most important action is to fix the wrong thinking as soon as possible.

The key Aristotelian insight, that observable things are worth studying as they are, may be applied in extreme fashion to teaching. According to an excessively Aristotelian view of studying student thinking, we physics instructors might be happy each day to ask our students questions, listen to their answers in detail, write down some observations, and thank the students for their time before inviting them to come back tomorrow. Later we would return to our notes and try to figure out what deeper truth was expressed in all those instances of student expression. But at no point would we distract ourselves from our scientific task with thoughts of what all that student thinking ought to have been. It is easy to see that such an extreme view is not appropriate for a classroom, because it leaves no place for student learning. Thus, neither philosopher's approach to nature is sufficient if it is applied to teaching, for in one, the reality of student thinking is neglected, and in the other, there is no role for instruction. If attention to student thinking is necessary for effective instruction, then the teacher is performing a complex, hybrid intellectual task of understanding clearly both where the students are, and where the teacher wants them to be. To prepare physics instructors is to prepare them to perform all the components of this hybrid task. We believe the greater task for those who prepare physics instructors lies in helping the instructors to understand the Aristotelian component.

Fortunately, however, establishing this Aristotelian component in the understanding of the instructor is not only a task of great effort, but also a great gift to the instructor. Through the recognition that teaching includes the work of making observations about student thinking and making sense of those observations, the instructor's vision opens into an entirely new realm of phenomena of infinite variety. And a scientist takes pleasure in observing and thinking about wonderful new things, as they are.

### A teacher solves puzzles of student thinking and learning

As if the news about a fascinating new class of phenomena weren't good enough, there is even more: the instructor also has the pleasure of *figuring out* puzzles of transforming student thinking from state A to state B, and then to state C, *etc.* Thus, teaching has a natural science component (understanding student thinking) and an engineering component (shaping student thinking to become more functional). These puzzles are as numerous as the instructor wishes, and include at least the number of students multiplied by the number of specific learning goals set forth by the instructor. A template for such a multi-part puzzle might look like "How can I better understand how [student name] thinks about [specific topic]? Once I understand well enough, how can I get [student name] to understand [specific aspect of specific topic]? How will I know if it worked?"

Kuhn described how it was this sort of intellectual pleasure that keeps scientists wanting to come back to work each day (Kuhn, 1962):

> *The challenge of the puzzle is an important part of what drives [the scientist] on… What then challenges him is the conviction that, if only he is skillful enough, he will succeed in solving a puzzle that no one before has solved or solved so well.*

Surely these words apply equally well to the science teacher, provided that the science teacher conceives of the job as providing a steady stream (really, a fire hose) of this sort of puzzle material. The main obstacle to conceiving of the job in this way, we believe, is merely the fact that those talented and intelligent people who might love to do such a job are not aware that the possibility exists. The lack of awareness is due in part to how such an approach to teaching has been modeled for potential future instructors by so few, and the other part, perhaps, is due to the lack of explicit instruction in professional development about how to approach teaching in this way.

## Theoretical perspective: Resources and framing

While teachers are making observations about student thinking, they are also making sense of what they observe: that is, they are inevitably (and probably implicitly) forming theories of what knowledge is and how learning takes place. This epistemological change is at the heart of transformative professional development. We assume that the cognitive processes underlying epistemological change are similar, in some respects, to the cognitive processes underlying conceptual change.

### Resources vs. (mis)conceptions

We take a *resources-based* rather than the more common *conceptions-based* view of the prior knowledge students bring to bear in physical science (Hammer, 2000; Smith, diSessa, & Roschelle, 1993/1994). In this view, learners (whether they are students or beginning teachers) have ideas that are activated by particular situations. People use these activated resources to construct knowledge and guide their behavior. These ideas are not categorically wrong or right, but rather are appropriate or inappropriate for the particular situation (Hammer, Elby, Scherr, & Redish, 2005; Russ, Scherr, Hammer, & Mikeska, 2008). Such a framework provides an explanation for how novices become experts: they begin to use resources from other contexts, adding new ones and building up a more coherent structure of ideas (Smith et al., 1993/1994). Smith et al. characterized such a framework as one that "emphasizes knowledge refinement and reorganization, rather than replacement, as primary metaphors for learning." (p. 116).

To see how the distinction between a conceptions-based and a resources-based cognitive framework plays out in interpreting learner epistemologies, consider a hypothetical introductory physics student named Dan. Even when a tutorial tries to elicit his intuitive ideas, Dan answers in terms of remembered facts and equations. Why? According to the conceptions-based framework adopted by most epistemology researchers (Hofer & Pintrich, 1997; Schommer), Dan's behavior probably stems from an epistemological *belief* that physics knowledge comes from authority, or that physics knowledge consists largely of facts and formulas (Hammer, 1994). In that case, changing Dan's behavior would likely involve an arduous attempt to confront and replace his epistemological "misconceptions" with more productive beliefs.

The resources-based framework provides an alternative interpretation and instructional strategy. In our view, Dan's focus on facts and formulas probably arises not from a stable epistemological *belief*, but from the context-sensitive activation of finer-grained *epistemological resources.* We see evidence of these resources, including resources regarding the source of knowledge, even in small children. When asked how she knows what's for dinner, a child might respond "Daddy told me," reflecting the activation of the resource *knowledge as transmitted stuff* (Hammer & Elby, 2002). The same child, when asked how she knows Mommy got her a present, might reply, "I figured it out, 'cause it's my birthday and I saw her hiding something." This answer reflects the activation of the resource *knowledge as fabricated stuff*, corresponding to the view that knowledge is built up from "raw materials" such as prior knowledge about birthdays and observations of sneaky parental behavior. Along these same lines, we can interpret Dan's behavior as stemming from the inappropriate over-activation of resources such as *knowledge as transmitted stuff* – resources that are useful when learning state capitals, but not as useful when learning physics concepts.

According to this framework, we can change Dan's approach to tutorials without introducing new epistemological beliefs "from scratch." Instead, we can help him activate epistemological resources *he already possesses* and applies in other contexts. For instance, Dan may rely on his common-sense ideas when explaining wave phenomena he sees at the beach, or when having a classroom discussion about psychological (rather than physical) phenomena. As instructors, we would try to help Dan "find" those resources and activate them in physics class (Hammer & Elby, 2003).

## Epistemological framing

A learning context such as a lecture or a tutorial can trigger a locally coherent set of epistemological resources, a set that explicitly or implicitly answers the question, "What counts as knowledge and learning in this context?" We call such a set of resources an *epistemological frame*, acknowledging the sociolinguistics literature that helped us formulate this perspective (MacLachlan & Reid, 1994; Tannen, 1993). Examples include *remembering stuff from authority* and *intuitive sense-making*. When Dan frames an activity as *remembering stuff from authority*, he "believes" – in that moment – that knowledge comes from authority and that solving a problem is a matter of finding the right knowledge. But in another context, Dan could frame problem-solving as *intuitive sense-making* and act as if he believes knowledge is constructed from everyday ideas and experiences. In a resources framework, neither of those frames is Dan's "true belief," because he doesn't have a single, universal belief about the nature of problem-solving.

When we study (and try to affect) the epistemologies that novice instructors exhibit while learning and teaching, the distinction between the conceptions/beliefs and resources models – and particularly, the notion of frames – becomes especially important. Instead of assuming that instructors *have* stable epistemological beliefs that they apply consistently and universally, we assume that different contexts can trigger different epistemological frames. For instance, when an undergraduate LA we'll call Lisa does her own advanced physics homework, she may fall – consciously or unconsciously – into manipulating symbols without trying to make sense of the underlying physics. When addressing "brainteaser" physics problems posed in her LA pedagogy class, however, Lisa easily enters an *intuitive sense-making* frame; she tries to make sense of the qualitative physics concepts and doesn't want to be told the answer, realizing that jumping to the answer is detrimental to her learning process. In that context, Lisa "knows" that learning physics is largely a matter of making sense of it for yourself, rather than just hearing a clear explanation or manipulating symbols. Nonetheless, when Lisa got into an introductory physics classroom as an LA, she acted as if she thought her students' learning is mostly a matter of hearing clear explanations and manipulating symbols. Our point here is that instructors' epistemologies appear to display context dependencies best analyzed in terms of resources and frames rather than beliefs.

Fostering epistemological development, in this framework, consists largely of helping novice instructors become more reflective about which frames they're using in which contexts (including when they teach), and eventually to "crystallize" their productive epistemological frames into well-thought-out epistemological beliefs about how people learn (Hammer et al., 2005). With Lisa, for example, we might try to create contexts in which she can observe students engaged in making sense of physics concepts, reflect on the similarities to (some of) her own most satisfying learning experiences, and consider how she might foster such learning experiences in the classrooms in which she is an instructor. This approach embeds our philosophical perspective: By posing student thinking about physics as an observable phenomenon, we support novice teachers in strengthening the Aristotelian component of their understanding of teaching Excitingly, we sometimes find that "once the phenomena [of student thinking] are adequately apprehended," novice instructors spontaneously draw valuable conclusions about cognition. Lisa, having thoughtfully observed students express a variety of ideas about force and motion, might express a new conviction that people hold multiple, perhaps contradictory ideas at the same time, or that people hold ideas that are neither right nor wrong but rather are applicable in certain contexts.

Our framework embeds a deep respect for novice instructors' existing beliefs and prior experiences, seeing them as the essential material from which expert conceptions of teaching are constructed The physics education community has long taken this perspective regarding learners' physics ideas, with the benefit that we can help students identify ideas that can be the basis for effective constructivist instruction (Hammer & Elby, 2003). Our theoretical framework extends this fruitful perspective to professional development We also go one step further by explicitly entrusting novices with the cognitive ability to figure things out: not only do they possess the building blocks of cognition, but also, we believe, they are capable of doing the building. This is in contrast to a view in which students need to be led step-by-step through a logical sequence prescribed by an expert.

## Methodology for facilitating and understanding transformative professional development

Our view is that professional development is most effective when it enriches the intellectual life of teachers and builds teacher identity by cultivating concern with student thinking. We are in the early stages of a project with two mutually supporting purposes: (1) to provide novice instructors with professional development that shapes what they value in science education, and (2) to better understand novice instructors' experiences of potentially transformative professional development.

### Design and implementation of transformative professional development

We have selected undergraduate Learning Assistants (LAs) as our initial target population because Learning Assistant programs can be centerpieces of university science education reform. LAs are talented undergraduates who work with faculty members to make large-enrollment courses more collaborative, student-centered, and interactive. Research has shown that LA programs improve undergraduate performance in physics courses, facilitate multi-disciplinary collaboration among faculty, involve more faculty in teacher preparation efforts, and recruit talented science majors to teaching careers (Otero et al., 2006; Pollock & Finkelstein, 2008). That is, a quality LA program is at once a mechanism for course transformation, for teacher recruitment, and for TA and faculty professional development.

LA programs integrate content, pedagogy, and practice and produce documented improvement in all three areas (Otero et al., 2006; Pollock & Finkelstein, 2008). LA programs are effective partly because LAs take a low-credit pedagogy course that instructs them in the nature of teaching and learning and introduces them to interactive teaching techniques. (TAs are usually not required to take a pedagogy class, partly because their typical course load is already very high.) Thus, in an LA program, university educators have the opportunity to shape participants' perceptions of teaching. LAs whose pedagogy class helps them appreciate the worth of reform instruction are a valuable resource for their institution and for science education reform generally.

Our design and implementation work is focused on the most flexible and most potentially innovative component of the LA program, which is the pedagogy session. The over-arching goal of these sessions is to provide LAs with a sense of the intellectual depth that a physics teacher can experience and enjoy by thinking deeply about student thinking. One promising activity that we have piloted is the Physics Interview Project (PIP). In this key example of an activity that originates in our theoretical perspective, participants interview peers about physics concepts and conduct structured reflections on the interviewing experience The activity holds promise for improving participants' abilities to listen and to teach effectively, increasing their esteem for the intelligence that teaching requires, and dramatically increasing their interest in K-12 physical science teaching as a profession.

In the PIP, each LA interviews a peer in order to study that person's thought process about some physical system. LAs perform three interviews throughout the academic year; the protocol for the first one is designed by the LA program director (HGC), and the later protocols are designed by individual LAs. After conducting each interview, LAs transcribe part of it and perform it in class, explaining why they thought the selection was interesting. They write a reflection paper that includes a characterization of the interviewee's thinking and the LA's thoughts on the difficulty and value of the experience. The PIP puts LAs face to face (literally) with the phenomena that are the subject of physics education research, and thus provides a particularly compelling entry point into that research. Importantly, LAs are exposed not only to the *refined results* of physics education research (as they might be in other LA programs) but also to the *raw data.*

## Research on participants' professional development experiences

The research arm of the project documents and analyzes individual instructors as they work together in preparation for teaching, as they teach, and in reflective interviews, with the goal of understanding transformative experiences they may have and learning about the processes by which novice instructors develop into intelligent and talented educators.

### Data sources

The aim of data collection for this project is to enable rich representation of individual LAs' experiences in the course of the program. To this end, we document novice instructors in all three environments that directly relate to their teaching: their pedagogy course (1.5 hours biweekly), their preparation course (1.5 hours weekly), and the classes in which they are instructors (2-4 hours weekly for each focal LA). Because the nature of the activity that people perceive themselves to be engaged in affects how they interact (Bateson, 1972; Goffman, 1986; Goodwin, 2000; MacLachlan & Reid, 1994; Tannen, 1993), we keep the recording subordinate to normal classroom practices. The approach has served us well in other projects (Goertzen et al., 2009; Goertzen, Scherr, & Elby, accepted (2010)).

In addition to the naturalistic videotaping, we conduct twice-yearly clinical interviews with a protocol similar to the one described in Goertzen et al, 2009, in which the primary question is "How is [name of course] going?" Open-ended prompts such as this have been shown to initiate conversations that result in epistemologically rich responses; participants volunteer which aspects of the course they judge to be successful, which aspects are lacking, what features they believe should get more emphasis, and otherwise reveal their values and expectations for instruction and their views about how students learn (Hammer, 1994).

### Analytic framework

The primary research activity of this project is to develop rich case studies of individual LAs' experiences in our professional development program, with the goal of understanding transformative experiences they may have. Case studies are particularly powerful for this project in that they get at the *mechanisms* by which our program may affect LAs. Fortunately, the increasing ease of video recording offers new opportunities to create richly detailed records of classroom activities. These recordings call for research methodologies that balance generalizability with interpretive validity.

*Classroom behavior as the primary data source.* When developing a case study, we use video recordings of LAs' classes as the primary source of data. We watch video episodes of LAs teaching and/or working together in the pedagogy and preparation courses and seek to provide plausible framings that might explain their classroom behavior. Transcriptions of the interviews are used to support or contradict the conclusions we develop while watching video episodes. This contrasts with many studies of TAs and teachers, which attempt to first understand the instructor through interviews or surveys, and then (in some cases) compare these assessments to actual behavior (Hammrich, 2001; Ishikawa et al., 2001; Lawrenz, 1992). We consider interviews to be secondary data because we conduct them with the intent of understanding LAs' overall beliefs and attitudes about tutorials, and thus they are unlikely to reference particular teaching situations (Goertzen et al., 2009). Depending on such interviews to fully explain TAs' framing would be attributing more consistency within an individual than we have argued is warranted.

*Insight-oriented analysis.* The primary purpose of "insight-oriented" analysis is to identify and richly describe the events in an episode as they unfold in the interaction among the participants. "Coding-oriented" analysis, in contrast, aims to generate categories that will be used to document the frequency and distribution of events of interest. (Of course, insight-oriented analysis may generate a coding scheme, and coding may lead to deeper insights.) The backbone of video analysis as we conduct it is collaborative analysis of short episodes of classroom activity. Having selected an episode, we play the video for a multidisciplinary analysis group, and members of the group call to stop the playback whenever they see something they want to discuss. We often benefit from replaying the video multiple times. In many cases, we do not approach the data with pre-conceived categories of interest. Rather, the goal of the group work is to progressively deepen our understanding of the participants' activities and to challenge the biases of individual analysts. There is a constant effort to ground assertions in the evidence of the video episode, and to limit proposed hypotheses to those for which the video episode (or others like it) could provide confirming or disconfirming evidence. The resulting discussion is often a lively, multidimensional conversation reflecting the diverse interests of the participants as well as the focal interest of the project.

### Observational methodology

The challenge presented by our theoretical perspective is to identify the ways that LAs frame their teaching and learning. No meter measures framing; instead, we infer people's sense of the nature of their activity primarily through their behavior, and secondarily through explicit statements in interviews. Participants' understanding of the nature of the activity in which they are engaged — i.e. their framing of the activity — guides their selective attention, provides cognitive structure for interpreting events, and manifests itself in their observable behavior.(Hammer et al., 2005) To the extent that framing is an interpretation based on previous experience, it is informed by an individual's broad history and experience with related events and systems. In the moment, though, participants mutually construct their sense of shared activity by means of verbal and nonverbal interactions, including linguistic signals, prosodic features, and body language (Bateson, 1972).

When we analyze LAs' teaching and learning behaviors, we use evidence such as how much they talk, the types of questions they ask, the conversational pace, their body positioning, gestures, and register (word choice, syntax, pitch, etc.), to infer how they are framing the situation. We look for additional support for these analyses from the ways LAs reflect on their teaching and learning in interviews. While we use LAs' stated beliefs to corroborate our ideas about how they frame, we are careful not to assume that these beliefs will necessarily match their actions. We also do not attribute framing solely to the LA: negotiations with students either support or undermine the instructor's sense of what's going on, so that together they construct a shared framing of the activity. (This is not to say that participants always have the same framing: mismatched framing is common, and can lead to humor or conflict depending on whether the participants recognize that they are framing in different ways (Goffman, 1986).)

## Findings

The main goal of our present work is to illustrate the fruitfulness of our theoretical and philosophical perspective rather than to demonstrate empirical results. That said, preliminary observations of the LAs' experiences with our professional development program include:

- LAs initially find it very difficult to ask questions to learn about another person's thinking. Their urge to guide them to a particular conclusion is so strong as to almost preclude thoughtful listening.
- LAs grow to understand the superficiality of questioning about vocabulary and trusting in answers in which proper vocabulary conceals the quality of understanding.
- LAs learn that people do not necessarily say what they really think, even if they are not trying to deceive.
- LAs come to recognize that the complexity of a real person's thinking is much greater than is indicated in summaries of research about tendencies in student thought.
- LAs learn that people who have no formal education in physics often have some surprisingly productive intuitions.
- LAs find that people can make a surprising amount of intellectual progress simply by having the opportunity to reason things out.
- LAs find the PIP valuable for relating to people in general, because of the practice with listening for understanding without imposing your own point of view.

The PIP seems to us to be extremely promising for transformative professional development. We look forward to learning more about the details of what happens in the interviews that the LAs conduct, and to better understand how the experience is meaningful to the LAs.

## References


Aristotle. (c. 350 B.C.). Prior Analytics. In R. Smith (Ed.). Indianapolis, IN: Hackett Publishing.

Bateson, G. (1972). A theory of play and fantasy. In G. Bateson (Ed.), *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology* (pp. 177-193). New York: Ballantine.

Finkelstein, N., & Pollock, S. (2005). Replicating and understanding successful innovations: Implementing tutorials in introductory physics. *Physical Review - Special Topics: Physics Education Research, 1*(010101), 1-13.

Gilreath, J., & Slater, T. (1994). Training graduate teaching assistants to be better undergraduate physics educators. *Physics Education, 29*, 200-203.

Goertzen, R. M., Scherr, R. E., & Elby, A. (2009). Accounting for tutorial teaching assistants' buy-in to reform instruction. *Physical Review - Special Topics: Physics Education Research, 5*, 020109 020101-020120.

Goertzen, R. M., Scherr, R. E., & Elby, A. (accepted (2010)). Tutorial teaching assistants in the classroom: Similar teaching behaviors are supported by varied beliefs about teaching and learning. *Physical Review - Special Topics: Physics Education Research.*

Goffman, E. (1986). *Frame analysis: An essay on the organization of experience*. Boston: Northeastern University Press.
Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489-1522.
Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction, 12*(2), 151-183.
Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics, 68*, S52-S59.
Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 169-190). Mahwah: Erlbaum.
Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *Journal of the Learning Sciences, 12*(1), 53-90.
Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89-120). Greenwich, CT: Information Age Publishing.
Hammrich, P. (2001). Preparing graduate teaching assistants to assist biology faculty. *Journal of Science Teacher Education, 12*(1), 67-82.
Henderson, C., Yerushalmi, E., Kuo, V. H., Heller, K., & Heller, P. (2007). Physics faculty beliefs and values about the teaching and learning of problem solving. II. Procedures for measurement and analysis. *Physical Review Special Topics - Physics Education Research, 3*(2), 020110.
Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88-140.
Hollar, K., Carlson, V., & Spencer, P. (2000). 1+ 1= 3: Unanticipated Benefits of a Co-Facilitation Model for Training Teaching Assistants. *Journal of Graduate Teaching Assistant Development, 7*(3), 173-181.
Ishikawa, C., Potter, W., & Davis, W. (2001). Beyond This Week. *Journal of Graduate Teaching Assistant Development, 8*(3), 133-138.
Kuhn, T. (1962). *The Structure of Scientific Revolutions*. London: University of Chicago Press.
Lawrenz, F. H., P; Keith, R; Heller, K. (1992). Training the teaching assistant. *Journal of College Science Teaching, 22*, 106-109.
Laws, P. W. (1999). New approaches to science and mathematics teaching at liberal arts colleges. *Daedalus, 128*(1), 217-240.
MacLachlan, G. L., & Reid, I. (1994). *Framing and interpretation*. Victoria: Carlton.
McDermott, L. C., & Redish, E. F. (1999). Resource letter PER-1: Physics education research. *American Journal of Physics, 67*(9), 755-767.
McDermott, L. C., Shaffer, P. S., & PEG. (1998). *Tutorials in Introductory Physics*. Upper Saddle River, NJ: Prentice-Hall.
McGivney-Burelle, J., DeFranco, T., Vinsonhaler, C., & Santucci, K. (2001). Building Bridges: Improving the Teaching Practices of TAs in the Mathematics Department. *Journal of Graduate Teaching Assistant Development, 8*(2), 55-63.
Otero, V., Finkelstein, N., McCray, R., & Pollock, S. (2006). Who Is Responsible for Preparing Science Teachers? *Science and Education, 313*(5786), 445-446.
Pollock, S. J., & Finkelstein, N. D. (2008). Sustaining educational reforms in introductory physics. *Physical Review Special Topics - Physics Education Research, 4*(1), 010110.
Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education, 92*(3), 499-525.
Schommer, M. (1998). The influence of age and education on epistemological beliefs. *British Journal of Educational Psychology, 1998*(68), 551-562.
Smith, J., diSessa, A., & Roschelle, J. (1993/1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3*(2), 115-163.
Sokoloff, D. R., & Thornton, R. K. (2001). *Interactive Lecture Demonstrations*. New York: John Wiley & Sons.
Tannen, D. (1993). *Framing in discourse*. New York: Oxford University Press.
Thornton, R. K., & Sokoloff, D. R. (1997). Using Interactive Lecture Demonstrations to create an active learning environment. *The Physics Teacher, 35*(9), 340-347.
Tobin, K., & McRobbie, C. (1997). Beliefs about the nature of science and the enacted science curriculum. *Science and Education, 6*(4), 355-371.

# Scaffolding Children's Understanding of the Fit Between Organisms and their Environment In the Context of the Practices of Science

**Kathleen E. Metz, UC Berkeley**
**Stephanie Sisk-Hilton, San Francisco State University**
**Eric Berson, Uyen Ly, UC Berkeley**

**Abstract:** This research project applies the learning progression perspective to the teaching of the conceptual underpinnings of evolution for 2nd and 3rd graders. We frame the progression from the epistemic perspective that understanding a scientific idea encompasses *using* that idea, in prediction, interpretation and explanation of the natural world. The progression foregrounds the question of the fit between organisms and their environment. This paper reports on the first step of the project, our teaching of the curriculum in an urban summer school and the analysis thereof. We analyze students' reasoning during prototype instructional activities and in comparison of pre- and post-interviews. We have had success scaffolding children's understanding of differential survival advantage of different traits. While we have existence proofs of children coming to understand the impact of this differential on changing distributions across generations, this next step of understanding appears to be much more difficult.

Evolution remains widely misunderstood at all ages. Preschoolers and adults alike manifest teleological reasoning in thinking about the biological, *albeit* with changing cues as to when it is applied (Keleman, 1999; Lombrozo & Carey, 2006). Across this same wide age-span individuals manifest 'biological essentialism", assuming that a species is determined by a defining essence (Gelman, 2003; Medin & Atran, 2004), while ignoring the within-species variation so crucial to understanding the mechanism of natural selection (Shtulman, 2005). Indeed even earlier biologists fell prey to these same misconceptions, as reflected in Ernst Mayr's (1988) tracking of teleological reasoning across the history of biology and Stephen Jay Gould's (1996) attribution of early evolutionary theorists' devaluing of within-species variability to their "essentialist tendencies".

In this continuity of challenges to understanding evolution, we envision an intriguing potential of starting instruction about the conceptual underpinnings of evolution at the second and third grade level. While these findings indicate that these conceptual challenges are robust and difficult to transcend for students of all ages, this continuity in the conceptual challenges leaves open the possibility that some form of early intervention might have an advantageous impact. We make no claim that these conceptual issues could be easily resolved if addressed much earlier in the K-12 curriculum. Rather, in accordance with the idea of *learning trajectory*, we posit that strategically taking up the conceptual underpinnings of evolution at this level could support a basic understanding of some facets of the theory and enable a deeper understanding of evolution at subsequent grades.

This research agenda accords with the new NRC report, *Taking science to school: Learning and teaching science in grades K-8* (Duschl, Schweingruber, & Shouse, 2007). This report emphasizes the fundamental importance of the interplay of instruction, experience, and maturation in the competence that children can achieve. Its conclusion that "*what children can do is in large part contingent on their prior opportunities to learn"* points to the key role of the learning trajectory in instructional design and analysis of children's capabilities. Finally in place of the problematic dichotomy of content and process, it recommends framing K-8 science learning in terms of four interrelated strands: *a)* Know, use and interpret scientific explanations of the natural world; *b)* Generate and evaluate scientific evidence and explanations; *c)* Understand the nature and development of scientific knowledge; and *d)* Participate productively in scientific practices and discourses. This research is grounded in this conceptualization of science learning.

This paper reports on our project, striving to scaffold young children's understanding of the fit between organisms and their environment. We present the instructional approach through description of the learning progression and design principles that informed the construction of the curriculum. We analyze the children's understanding of the fit between organisms and their environment and how this changed over time, through: *a)* comparison of pre- and post-test one-on-one interviews; and *b)* children's thinking wrestling with the targeted ideas in the context of prototype lessons. The research is small scale, limited to two cycles of research within a summer enrichment program (over which the team will have complete control of student placement and instruction) and two years of the classes of three second and third grade teachers in a diverse, urban school (reflecting the complex environment to which any viable intervention must adapt). This paper reports on the first cycle of the summer enrichment program.

## The Scaffolding of Children's Understanding

### The Learning Progression

Above and beyond consideration of where children begin and the long-term end-point in the conceptualization of a learning progression, we aimed to conceptualize the conceptual terrain of the progression such that it supports increasing explanatory power *from the students' perspective*. Most elementary school science curricula violate this criterion, with content that will only prove to have some explanatory power when linked to concepts reserved for older grades – a curricular pattern that fails to reflect the power or purpose of science or the heart of its practices. We aimed to construct the learning progression in evolution in such a way that the concepts within this grade-band have explanatory power in and of themselves, while also strategically grounding more powerful and complete explanatory models to be taught in subsequent grade levels.

The learning progression is framed in terms of *increasingly powerful explanations of the fit between organisms and their environment.* We focus on phenomenology and change at the time scale of microevolution. An ecological perspective is limited to that needed to understand this relationship; *e.g*., the survival value of different inheritable traits of a population within a particular environment. The cellular level is excluded, as well as speciation and genetics above and beyond the idea of inheritance.

Our current version of the learning progression consists of seven levels. We assume that children will come to school with some understanding of the two most basic levels: the simplistic idea that organism*s Live where they belong* (Level I) and the more adequate idea that organisms *Live where they can get what they need* in Level 2. While children come to school with some ideas about what organisms need, the curriculum aims to develop these understandings. Variation in organisms' structures/ limiting factors embodies the idea that organisms are able to get what they need in very different environments --- with different limiting factors --- due to variation in their structures. At this level we can build on some understandings that children have of structure/ function, but also need to substantially elaborate the structure/ function idea and the scope of the context to which it is applied. *Survival value of specie's trait* (Level 4) conceptualizes fit between organisms and environment in terms of the survival value of a particular trait (e.g.; the survival value of the Grey Whales' migration between Baja and the Arctic or the ridge on the male crickets' wing that enable them to chirp). Whereas survival value or cost-benefit analysis may be intuitive at some level and some contexts (*e.g.*; is it worth trying to get my ball back now from the playground bully), we doubt children have ever considered animal or plants' traits from this perspective. *Within-population variation and differential survival advantage* (Level 5) takes up the idea of differential survival advantage of different traits of the same characteristic within a given environment. *Natural selection* (Level 6) is conceptualized as the idea that, over many generations, inherited traits that help organisms' chances to survive and reproduce in that environment become more common there. Traits that hurt their chances become less common. The traits that hurt its chances become less common. The last level (Level 7) encompasses the outcome of the mechanism of natural selection, namely *Organisms well-adapted to where they live.* A secondary progression, supporting understanding of life cycle, resemblance of parent and offspring, and inheritance supports the primary progression building increasingly powerful explanations of the fit between organisms and their environment.

### Pedagogical Design Principles

Given space constraints, we present the pedagogical design principles here in brief:

1) Understanding a scientific concept entails *using* the concept in the practices of science, including interpretation of the natural world, making prediction, and developing explanations.
2) Build children's conceptual understanding through a range of scientific knowledge-building practices, including thought experiments, field-based and laboratory-based empirical inquiry, and text-based research.
3) Immerse children in exploration of a phenomenology and the puzzling patterns therein *prior* to introducing the corresponding explanatory abstraction.
4) Leverage strategically selected in-depth cases of phenomenology and their interpretation as a basis to build generalizations and abstractions.
5) Capitalize on fruitful preconceptions.
6) Build from contexts that we anticipate will be less likely to evoke buggy reasoning before having students apply these new ideas to contexts that we anticipate will be vulnerable to buggy reasoning.

7) Emphasize the metaconceptual and metacognitive knowledge that support children's understanding of the power of the targeted ideas in explaining the fit between organisms and their environment.

### The Curriculum Modules

We have developed two curriculum modules, to scaffold children's advancement on the learning progression, in accordance with these design principles. One of the modules develops these ideas in the children's study of botany, the other in the realm of the children's study of the animals and their behavior. Each module is approximately 30 hours in length, about the length of a relatively long "replacement unit".

## Measures of Student Understanding

### Pre- and Post-Interviews

We developed a structured interview instrument to measure student progress on the learning trajectory. In accordance with our conceptualization of what understanding of a concept entails, items elicited children's using the targeted ideas in predicting, interpreting and explaining biological phenomena. The instrument consisted of 7 extended items. One item was framed in terms of one of the particular organisms the class studied (Brassica rapa on the part of the botany class and crickets for those in the animal class). All the others were transfer items, a characteristic of the instrument that we realized made it extremely difficult. The instrument included items without any kind of scaffolding, as well as two items with dynamic scaffolding and two items with what we conceptualized as empirical scaffolding, in the form of empirical feedback following the children's generation of a prediction and the explanation thereof. The instrument assessed the conceptual terrain through: *a)* two items involving children's thinking about why particular organisms (otters and kelp) lived where they lived and consideration of where else they could or could not live; *b)* four cases of microevolution, involving predictions and explanations about a population after an environment change, followed by explanation of what actually happened (*e.g.*; changes in the coloration of male guppies, following arrival of predators); and *c)* explanation of change in a characteristic over time (e.g.; the question, drawn from this research literature, of why cheetahs today are so much faster than their ancestors).

### On-Line Student Negotiation of Curricular Activities

Above and beyond analysis of gains in understanding from before and after the instructional intervention, we are also interested in understanding how the curriculum functioned to support conceptual development and how we might iteratively improve it for these purposes. For the purpose of analyzing the interplay of instruction and learning, we videotaped all instruction and developed a complete database of student written work and easel pad record notations from class discussions (used in place of the white board to support the inclusion in the written artifact data base.) We also designed embedded assessments into the curricular plan.

## Analysis of Children's Understanding

We conducted structured one-on-one interviews with almost all of the 40 children in the summer program. These interviews were videotaped to enable close analysis. We are currently formalizing the coding process to apply to the full set of interviews. While we will report statistics about all 40 children and their changes from pre- to post at the conference, the results we report are only preliminary and formative, and based on the 13interviews we have studied to date. (These * are all the children from the animal behavior class who participated in all days of the program, minus one child with Down's syndrome who did not participate in the post-test interview.

### Advancements From Pre- to Post-Test and Enduring Conceptual Challenges

Even before the curriculum, the majority of the children we have analyzed (9 out of 13) had some understanding of the differential survival value of some traits, at least in the relatively straightforward and familiar instance of camouflage; *i.e*., organisms that stick out in an environment with a predator, as opposed to those that blend in, are more likely to be eaten. This in-coming level of understanding emerged in our framing of an item based on the classic case of the shifts in peppered moth coloration following severe pollution due to industrialization, as well as another item based on Endler's (1980) study of microevolution of male guppy coloration as a function of the presence of predator fish. For example, in an item asking children to make predictions and explanations thereof of the color of moths following heavy soot darkening their landing places, most of the children realized the survival value of dark coloration. For example: "They [the birds] can find the light ones because they're not camouflaged. But the dark ones, they're camouflaged. But they won't find the dark ones."

In the single non-transfer item, 12 of 13 children reflected some degree of conceptual advancement from pre to post-test. Thus, for example, the animal curriculum cohort had studied a case of microevolution of crickets on Kauai (Tinghitella, 2008, 2009): the appearance in the island cricket population of a slightly different wing structure such that cricket could not make a chirping sound, followed by arrival of maggots that lay their eggs in crickets that they can locate through their chirping. Post-instructional intervention, 92 % of the children in the animal curriculum cohort analyzed to date not only appreciated the differential survival advantage for the particular population currently living *( i.e;* those crickets that chirped were more likely to be found by the maggots and thereafter soon die from the development of the maggots in their body), but also considered the impact of this differential survival value -- combined with some rudimentary understanding of inheritance-- on the relative frequency of the trait in the next generation. For example, Ellen appropriately predicts an increase in the proportion of non-chirpers, but qualifies her prediction with noting that chirping is both a risk and an advantage. On the one hand, she views chirping as a survival disadvantage, in her words, "The fly will only go for the chirping cricket". On the other hand, she also notes the survival value of chirping, "They [indicating male crickets with changed wing structure] don't chirp, so that mates won't be attracted to them." As she explains her prediction of an increase in non-chirpers and decrease in chirpers (as represented in iconic symbols for chirpers and non-chirpers in the next generation), "a lot of the chirping crickets got eaten by the fly, the fly. And then most of these guys [non-chirpers] gave some birth to offspring."

Analysis of the transfer items revealed that 69% of the children also considered the impact of the shifts in survival value of traits to changes in the relative frequency of the traits in subsequent generations to contexts they had not studied in class. For example, consider Wally's post-test reasoning about the case of microevolution of male guppy coloration post arrival of predator fish. He predicts there will be more "black" or "grey ones" (camouflaged) and fewer "yellow ones" (colorful), on the grounds that:

> The black guppies can blend in more than the golden guppies... The yellow ones are getting eaten, because they can't blend in so good. And the gray ones... most of them, they can survive and then they give babies. There will be more and more gray ones... The yellow guppies, they'll get eaten so when they give babies, they won't have as much, and when they get eaten, there will be less and less.

In short, these interviews documented second and third graders' wrestling with questions of microevolution. The post-instruction interviews revealed conceptual advancements on the part of many students in a context they had studied, success on transfer tasks on the part of some, as well as multiple enduring conceptual challenges. In the sections below, we consider the functioning of the curriculum vis à vis these challenges.

## On-Line Student Negotiation of Curricular Activities and Embedded Assessments

### Children's Thinking in the Context of the Botany Curriculum

The botany curriculum included an instructional sequence of an empirical investigation followed by a thought experiment designed to support the children's wrestling with several ideas central to the learning progression, including within-kind variation (LP5), differential survival advantage (LP5) and a change in the prevalence of traits in subsequent generations (LP6). These investigations engage students in applying these ideas in the context of making predictions and developing explanations. Students are also immersed in a phenomenology to establish the basis from which to reason scientifically. The investigations build on prior investigations to develop these abstract ideas in a phenomenology that is familiar to students.

This instructional sequence begins with the students' close analysis of a population of aster seeds, identifying subtle differences between seeds. Students noticed that some seeds were slightly bigger than others and some had slightly bigger parachutes. They made predictions about which of these differences might matter for the distance that the seeds might travel in the wind. Students developed an experiment to test their predictions, comparing the distance bigger seeds and smaller seeds traveled.

To extend their empirical exploration of traits that support dispersal potential, we designed a thought experiment to scaffold students thinking around differential survival value and subsequent generations. We adapted the thought experiment from a case in the research literature of microevolution of the aster seed structure following the arrival of the seeds on ocean islands (Cody & Overton, 1996). As part of the thought experiment, we asked students to make predictions about the survival advantage of different aster seeds on the island and the distribution of size within the population of aster seeds in future generations. In this limited space, we focus on the interplay of instruction and evidence of students' thinking along the learning progression as reflected in the thought experiment.

We framed the microevolution thought experiment in terms of a population of aster seeds (large and small) that landed on an island. In this thought experiment, students (a) made predictions about where the next generation of seeds might land; (b) explained their predictions in terms of the survival advantage of the seeds that land on the island; (c) made predictions about future generations of seeds on the island and their differences with respect to size; and (d) compared their predictions to empirical data from the scientist. The students worked together as a class on each part of the thought experiment, but recorded their own predictions and explanations separately on written templates functioning as independent, instructionally embedded assessment. An analysis of the classroom video of the thought experiment and written assessments reveal that most children had some understanding of survival advantage of the larger seeds and that the larger seeds would be more prevalent in the next generation (LP6).

Consider the interplay of task and student thinking in the context of this thought experiment. Shown a sample of seeds that landed on the island, the teacher challenged the class to make predictions about where the next generation of seeds might land (Figure 1). As part of the classroom discussion, many students shared similar predictions. For instance, "What I think will happen is that most of the heavy seeds will land on land and the light seeds will likely land on water" and "What I think will happen is that the light seeds will be more likely to land in the water because they fly farther." Based on analysis of student work, all of the students represented more heavy than light seeds in their predictions of which seeds would land on the island in the next generation.

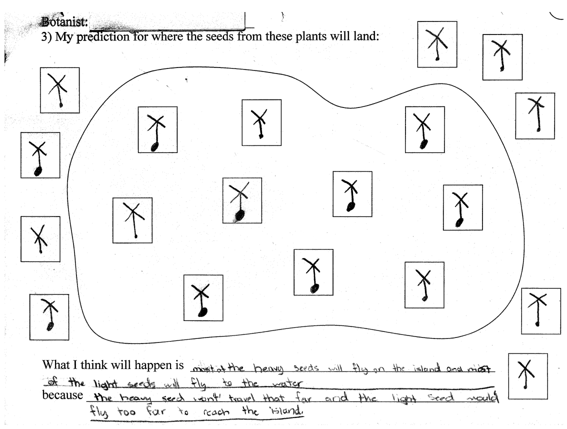



Figure 1. Sample of prediction of regarding where seeds with different traits will land.

The teacher asked the students to explain their prediction in terms of survival advantage, using a written template to scaffold their explanations. Analyses of these assessments reveals that 12 of 17 students appropriated the class discussion about the impact of seed size on distance traveled and used this as a rationale for explaining the differential survival advantage of heavier seeds. These students explained that the heavier seeds had a survival advantage relative to the lighter seeds. They reasoned that the lighter seeds tended to travel too far and would land in the water, whereas the heavier seeds didn't travel as far and would be more likely to land on the island, e.g.:

> [*Written prompt:* The heavy/light (circle one) seeds will have a survival advantage because…] …[circled "heavy" above]… the big ones will travel on land and the light ones will travel too far so they might fall in the water.

Students subsequently made predictions about what the scientist saw when she went to study the seeds on the island, many years after the seeds first arrived. Many students noted that while there will likely be MOSTLY bigger seeds on the island, there may still be some smaller seeds, for example:

> I predict that most of the…there won't be that much little ones that will be on the island. But I think the heavy ones will be on the island, because they have the advantage because they are heavier and they don't go as far…

In the final part of the thought experiment, the teacher presented the students with empirical data from the scientist and asked the students to explain what the scientist saw, how that might have happened and how their predictions matched what the scientist saw. Consider, for example, how one student compared her prediction to the empirical data (below). After summarizing the scientist's observations, the student explains these observations by reasoning that the lighter seeds were less plentiful because they had flown off of the island. She then concludes that her prediction fit with the scientists' observations based on a relative distribution of small to big seeds.

> *What did the scientist see?* A bunch of big seeds and a few little.
> *Why do you think that happened?* All the little seeds flew off the island.
> *How does your prediction fit with what the scientist saw?* That my prediction fit in because I drew three small seed and a lot of big seeds.

Note she appropriately predicted a decrease in light seeds and an increase in big ones, she may not fully attend to the generational nature of the mechanism over long periods of time.

Analysis of students' work indicates that 12 of 17 students demonstrated an understanding of survival advantage (LP5). However, generational thinking was still problematic for many. Nonetheless, by immersing students in the seed phenomenology and by engaging in the practices of science under these instructional conditions, students were able to reason about the differential survival value of traits and how the value can change with the change in the environment (arriving on the island.) The curriculum subsequently built on this emerging understanding to further develop student reasoning along the learning progression. In subsequent investigations, students continued to build on these ideas in a more independent context of partner work focused on using empirical data from their plant populations growing in the classroom to make predictions about plant characteristics in future generations following an environmental change.

### Children's Thinking in the Context of the Animals and Their Behavior Curriculum

The idea of multiple generations also proved particularly difficult for students as they reasoned about shifting survival advantage of a trait in relationship to an environmental press in the context of the animal behavior curriculum. In the Kauai crickets thought experiment, the first point at which the curriculum scaffolded students to work through the process of natural selection most students (88% based on analysis of individual students' worksheets)readily worked into their repertoire the idea that a particular trait could have *survival value* for an individual (LP 4), and also that within a population, individuals with a certain trait could have a *survival advantage* over those members of the population without the trait (LP 5). Likewise the idea that offspring usually had traits similar to the parents was non-problematic according to students' comments in the class discussion and as shown universally on student workseets (although students over-attributed identical traits to offspring, since the curriculum did not explore genetics). Some students were able to use these ideas to reason about the likelihood of a trait being passed from one generation to a single subsequent generation. For instance, at the beginning of the Kauai Crickets thought experiment, Maria explained the ideas of her partner, Carlos, regarding which type of male crickets had a survival advantage: "the chirping ones because um, the chirping ones could find a mate, and then when they had their babies they could also chirp." However, almost all students encountered significant difficulty in reasoning about the impact of an environmental press over many generations. In their written records, 23% predicted that variation would simply cease to exist in the next generation (the trait with survival advantage would be present in all individuals), 29% predicted an increase in the population of non-chirpers but were unable to describe their reasoning, and 23% were unable to construct a response at all. Of the 23% who predicted an increase in the population of non-chirpers and gave a reason, all focused on the fate of an individual or a single generation, for example, "the chirping crickets get caught by the fly." Students had difficulty creating a chain of reasoning that moved beyond the second generation and considered the impact of differential survival value *survival, reproduction, and traits present in subsequent generations.*

In the animal behavior curriculum, a single external representation used multiple times throughout the series of investigations appeared critical in scaffolding children's multi-generational reasoning (see figure 2). Evidence suggests that this form, in the context of the curriculum's thought experiments, helped students develop evidence-based predictions about the shifting distribution of a trait over many generations. The representation was first used in the Kauai crickets thought experiment, where it did not have an obvious impact (see above). However, in subsequent thought experiments, when students returned to the representation they were better able to make generational predictions. For instance, in students' final projects, in which student pairs predicted changes to a population of an animal of their choosing in response to a real environmental press, the teacher observed only 2 pairs reasoning about a potential shift in trait distribution due to differential survival based on traits. Upon bringing back a version of the generational chart seen in figure 2, all pairs were able to construct such an explanation (with differing degrees of direct teacher intervention). The generational representation scaffolded students' development of a multi-generational prediction regarding shift in the distribution of a traitin three key ways, discussed briefly below in the context of the Kauai crickets example.

First, the representation *required students to represent the members of a population as individuals with or without a key trait and on this basis make predictions about whether they were likely to survive and reproduce.* After students discussed the survival value of chirping for male crickets prior to the arrival of a predator, the teacher presented a small sample population to represent the initial trait distribution, in this case 10 male crickets in which 7 were chirpers and 3 were non-chirpers. The teacher then described the arrival of the predator fly, and several students co-constructed the idea that non-chirpers, previously at a disadvantage in attracting a mate, now had a survival advantage because they would be less likely to be victims of the fly. However, the seemingly simple question of "would there be more or fewer chirping crickets in the next generation" met with little response. The teacher then asked students to predict using actual numbers: out of the 10 crickets represented on their page, which five would survive and reproduce. Underneath the first representation were 5 boxes in which to place the reproducers. They then filled in connected boxes to show the traits of their offspring. Students repeated this process for two subsequent generations before making predictions about the population 15 generations later. Manually "moving" representations of individuals from members of the population as a whole into a sub-group of those that reproduced allowed students who were unsure of the question when initially asked to develop well-reasoned responses. For instance, Tyrone, who adamantly insisted that neither chirpers nor non-chirpers had a survival advantage when the predator fly arrived, when forced to make a decision about which individuals to place in the "reproducer" boxes, placed more non-chirpers than chirpers, and wrote [italics refer to provided prompt]: "*Now I think* the non-chirping *male crickets have a survival advantage because* the fly would not eat them." In class conversation and on his representation sheet, he contended that 15 generations later, there would be *no* chirpers. While this is not entirely accurate, his reasoning showed a move from not accounting for survival advantage to predicting individual survival based on a trait to making predictions about trait distribution in the population as a whole.

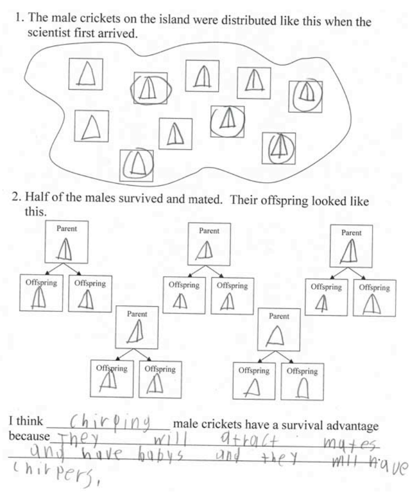



Figure 2. Sample of simplified visual representation of shifting distribution of a trait over generations.

Second, this format provided a simplified representation of *next generation as involving the offspring of organisms present at the onset of the investigation.* The representation physically linked traits of offspring to traits of parents. Students then transferred only the offspring into a new box labeled "Generation 2." Of course, this is a simplification in many ways. The representation limited offspring per parent, mandated number of reproducers in a generation, and assumed offspring always had the same trait as the represented parent. Most importantly, the representation prompted students to reason about the fate of individuals, although the goal was to get them to reason about the shifting distribution of a trait within a population. However, for many students, as in the example of Tyrone above, reasoning at the level of the individual appeared to be a necessary step to move from the claim that it was impossible to predict future distribution toward reasoning that accounted for differential survival advantage resulting in some members of a population being more likely than others to survive and reproduce in each generation (LP 6).

Finally, the generational representation provided a simplified visual that scaffolded students' understanding of the *shifting relative concentration of a particular trait in a population across generations.* By looking at the scattered icons representing the predicted population by trait in each generation, students easily discussed the shift in prevalence of a trait over multiple subsequent generations. Some students described the shift using the actual numbers of the greatly simplified population drawing, for example, "here there's seven that chirp, but now there's just four so there's more that don't." Others were able to make more generalized statements about how traits become more or less prevalent across generations, still referring to the simplified representation. For instance, Martin stated, "[in generation 15] there'll be mostly non-chirpers because if they had to mate… and if they had babies, then they'd have to be non-chirpers." Many students continued to need this representation to scaffold their predictions and explanations for the remainder of the curriculum, even those who were able to construct more generalized explanations of shifting trait distribution. As discussed above, 7 of the 9 student pairs used the representation in constructing an explanation for their final project in the curriculum.

## Conclusions

The understanding of evolution poses challenges for high school and colleges students. Our project aims to support understanding of this key theoretical frame by beginning early: with second and third graders. We have conceptualized a learning progression in terms of steps of increasing power and complexity in explaining the fit between organisms and their environment. Analysis of structured interviews, classroom discussions, and instructionally embedded assessments reveal children wresting with key aspects of natural selection in cases of microevolution. Predicting changes in relative frequency of a trait from one generation to the next appears to pose less substantial challenge than extrapolating shifts in distributions across a broader scope of generations and time.

We have three more rounds of the educational design experiment, including a second summer school with many of the first summer's second grade participants as third graders in the other domain (switching from study of botany to animals and their behavior or visa versa) and two semesters in regular school year classrooms. This gives us the opportunity to continue to rework the curriculum to more adequately scaffold the conceptual issues that prove most difficult. Our goal is to iteratively refine the instruction, as we closely examine the power and limitations in children's understandings of evolution that emerge under these instructional conditions. This process will best position us to try to differentiate robust limitations of children this age from limitations due to suboptimal learning opportunities.

## References


Cody, M. L. & Overton, J. McC. (1996). Short-term evolution of reduced dispersal in island plant populations. The *The Journal of Ecology*. 84 (1). 53-61.

Duschl, R., Schweingruber, H.A, & Shouse, A. (2007). *Taking science to school: Learning and teaching in grades K-8.* Washington, DC: National Academy Press.

Endler, J. (1980). Natural selection on color patterns of Poecilia reticulata. *Evolution*, 34, 76-91.

Evans, E. M. (2000). The emergence of beliefs about the origin of species in school-age children. *Merrill-Palmer Quarterly*. 46 (2), 221-254.

Fay, A. L. & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding in indeterminacy. *Child Development*. 67 (2), 689-716.

Gelman, S. A. (2003). *The essential child*. Oxford UK: Oxford University Press.

Gould, S. J. (1996). *Full house: The spread of excellence from Plato to Darwin*. NY: Three Rivers Press.

Keleman, D. (1999). The scope of teleological thinking in preschool children. *Cognition*, 70, 241-272.

Lombrozo, T & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition.* 99, 167-204.

Mayr, E. (1988). *Toward a new philosophy of biology: Observations of a evolutionist*. Cambridge: Harvard University Press.

Medin, D. & Atran, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review*. 111 (4), 960-983.

Shtulman, A. (2005). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology*.

Tinghitella, R. M. (2008). Rapid evolutionary change in a sexual signal: Genetic control of the mutation 'flatwing' that renders male crickets mute (Teleogryllus oceanicus) mute. *Heredity*. 100, 261-267.

Zuk, M., Rotenberry, J. T., & Tinghitella, R. M. (2006). Silent night: Adaptive disappearance of a sexual signal in a parasitized population of field crickets. *Biology Letters.* 2, 521-524.



## Acknowledgements

This research was funded by the National Science Foundation Grant DRL-0814821. The opinions expressed herein are those of the authors and do not necessarily reflect those of the NSF.

# Learning physics as *coherently packaging* multiple sets of signs

Kristine Lund, Karine Bécu-Robinault ICAR Research Lab, CNRS, University of Lyon
ENS-LSH, 15 parvis René Descartes, 69007 Lyon, France
Email: Kristine.Lund@univ-lyon2.fr, Karine.Robinault@inrp.fr

**Abstract:** This paper studies the multimodal reformulations that teachers and students make when they talk about and do physics experiments in class. Using the framework of semiotic bundles, we show that reformulating aspects having to do with physics knowledge while moving between talk, gestures, drawings and manipulations is done differently by experts and novices. Analysis of video excerpts illustrate that teachers are able to coherently package multiple sets of signs throughout their discourse and actions in the classroom, but students who are learning physics have specific problems that this framework makes evident. In particular, successfully reformulating from one semiotic resource to another implies that the first resource be correctly constructed. In addition, specific *tool affordances* hinder students in their attempt to coherently package multiple sets of signs while this is not the case for teachers. We conclude by suggesting ways in which teachers can ease students' difficulties in constructing semiotic bundles.

## Introduction

Our research team has many years of experience in designing teaching sequences for upper secondary school in physics and chemistry (e.g. Buty, et al., 2004). These groups are based on the collaboration of practicing teachers and researchers in which instrumental and theory-oriented approaches are combined (Brown, 1992). One main hypothese stemming from research on the design of these teaching sequences is that the modeling activities of students in relation to an epistemological point of view concerning physics' functioning must be taken into account (Bécu-Robinault, 2002). In addition, other theoretical underpinnings are taken into account, such as the role of social interactions in learning (e.g. Doisy & Mugny, 1984) and research results concerning misconceptions (cf. Duit & Von Rhöneck, 1998 for a meta-review). Since 1997, we have developed a range of teaching-learning sequences, concerning learners aged from 15 to 18 all of which have been implemented in classrooms and evaluated from teaching and learning perspectives. Our results show that teachers are able to use the teaching-learning sequences and associated documents and they also perceive improvements in student learning. Despite this, evaluations also show that, unexpectedly, a variety of teachers devote a great deal of time to the reformulation of the ideas in the elaborated documents (Lund & Bécu-Robinault, forthcoming). We know teachers have a fixed duration to teach their class and do not purposefully waste time, so we hypothesize that this phase is important for helping students grasp the knowledge to be taught. In this study we explore the implicit reasons for which teachers may perform such a reformulation activity and this prompts us to look more closely at the initial lessons of a particular physics teaching-learning sequence.

In what follows we will present our theoretical framework, describe our empirical study, present our analyses and results and conclude with perspectives for further work.

## Theoretical framework

Physics learning in the classroom is a complex activity that is both cognitive and social where teachers and students use talk, gestures, drawings and the manipulation of objects to co-construct physics concepts. It has been shown that embodiment through gesture plays a role in learning new concepts (Goldin-Meadow, et al, 2009) and Roth & Pozzer-Ardenghi (2005) propose that for understanding communication in everyday settings, one must take into account not only words and gestures but also all other semiotic resources co-participants produce or find in the setting. In this paper, we choose to address students' conceptual difficulties in the particular case of learning electricity through the study of multimodal reformulation as a tool to co-construct discourse (De Gaulmyn, 1987; Apotheloz, 2001; Lund, 2007), thus taking into account all the semiotic resources that co-participants do. We will use the notion of semiotic bundles as a method for explaining students' difficulties (Arzarello, 2004) and illustrating teachers' expertise. In the sections that follow, we present these notions and set the scene for showing how reformulating aspects of one particular semiotic resource into another one is an expert activity for teachers but fraught with difficulties for students.

### Multimodal reformulation

The term multimodal is often used to signify the medium in which a particular message can be expressed, for example text or graphics (e.g. Pineda & Garza, 2000). The authors Kress & Van Leeuwen, (2001) distinguish between modes and media: modes are the abstract, non–material resources of meaning-making whereas media are the specific material forms in which modes are carried out. The mode of gesture is carried out in the media of movements of the body. Different media afford different kinds of meaning (Dicks, et al. 2006), e.g. expressing an idea in writing or speech affects what is conveyed. Here, we use the term multimodal to describe the addition of non-verbal human face-to-face interactive phenomena such as gesture, gaze, posture, object manipulations, etc. to speech, studied extensively as a phenomenon in its own right by researchers such as Kendon (2004), Cosnier, (2000), McNeal (1992) and Brassac, et al. (2008).

Many interactive situations whose objective is learning presumes a dissymmetry of knowledge between interlocutors. This dissymmetry calls for adjusting discourse so that interlocutors reach mutual comprehension and common ground. Teachers' discourse does not escape from these adjustments and they often occur as reformulations; what learners say is put by the teacher into more conventional words (Chouinard & Clark, 2003). In this study, we borrow a different focus for the definition of reformulation, elaborated in the context of collaborative writing (Apotheloz, 2001). This is an oralo-graphic situation that articulates two different modes, speaking and writing. For Apotheloz, cooperation during such a task consists in "continuously exhibiting, that is at each step, the manner in which what is formulated articulates with what has already been formulated" (p. 62, our translation from French). We study this same phenomenon of reformulation, but in a situation where there is a plurality of multimodal activity: teachers and students speak, write, gesture, draw and manipulate objects and we see that the reformulations between modes that are carried out by teachers and those that are carried out by students differ greatly in how such a reformulation provides for the construction of meaning.

### Semiotic Bundles

We use Arzarello's (2004) semiotic bundle to interpret such multimodal reformulation, originally defined to analyze interactions around mathematics learning. A semiotic bundle is a collection of semiotic sets and a set of relationships between the sets of the bundle. A semiotic set is composed of three elements. The first component is a set of signs that may *be produced* with different intentional actions (speaking, writing, drawing, gesturing, handling an artifact). The second component is a set of modes *for producing* the signs and possibly transforming them. The modes can be rules or algorithms, but can also be more flexible action or productions modes, such as the modes referred to in the previous section, comparable to the intentional actions of Arzarello (speaking, writing, drawing, gesturing, handling an artifact). The third and final component is a set of relationships among the produced signs and their meanings embodied in an underlying meaning structure. A semiotic bundle is a dynamic structure changing over time due to the semiotic activities of the participants who are constructing it.

Our research questions focus on describing how expert teachers and novice students reformulate aspects having to do with physics knowledge while moving between talk, gestures, drawings and manipulations of physical objects. We hypothesize that experts and novices will reformulate differently and that the theoretical concept of the semiotic bundle is useful for understanding students' difficulties in physics while illustrating how teachers' expertise in multimodal reformulation can potentially be leveraged for helping students.

## Empirical study

This study has been carried out in the context of a research-action group that designs teaching sequences for physics (Bécu-Robinault, 2007; Buty et al, 2004). All the teaching sequences have been co-elaborated by researchers and practicing teachers, implemented in classrooms and evaluated from teaching and learning perspectives. In this presentation, we focus on the electro kinetics sequence, lasting 3 months, for grade 7 in the French school system. The teaching and learning of electricity has been the object of many investigations in science education. From international research results, we know that students encounter deep-level conceptual and reasoning difficulties in understanding introductory electricity. The main difficulty concerns notions of current, electric current and energy that are not differentiated by students (Psillos et al., 1988). Many teachers spontaneously use analogies, such as a water analogy, to teach electricity. Indeed, using analogies in teaching is thought to provide learners with tools that facilitate science understanding and promote conceptual change. We thus chose to introduce an analogy, in order to predict and interpret phenomena without using any formal concepts and to help students to develop an understanding of electricity and energy concepts (cf. Scott et al., 2006). This analogy relates these concepts to familiar every day objects (loaves of bread represent energy, delivery vans in motion represent the current, the bakery and the supermarket are respectively the analogues of the battery and the light bulb). The analogy is intended to help students to differentiate the concepts of current and energy. The lesson is set up so that questioning why a battery has a life span prompts thinking in terms of energy while questioning how a bulb lights up prompts thinking in terms of current.

### Short description of the lessons studied

In the French curriculum, electricity at grade 7 is taught through a phenomenological study of electrical circuits. In this presentation, we will focus on the first two lessons dedicated to the study of the simple circuit (how to light a bulb with a battery) and to the introduction of an analogy used for teaching (differentiation of current and energy, although we will not focus on energy in our examples). The first lesson is a necessary preliminary, in order that students understand how to handle electrical devices, identify the different terminals, and have no choice but to build a loop with the battery, wires and bulb. To understand how a simple circuit functions, students are asked to draw their experiment before handling electrical devices. This lesson begins with the presentation of a well-known object, a Maglite™ flashlight. Students usually question a battery's freshness and this cannot be interpreted in terms of current so the second lesson addresses this concern. With the help of the analogy, we suggest that the current implies thinking with a circuit perspective and that energy implies thinking with a chain perspective. Students are thus asked to connect each term of the analogy and its corresponding physical object in the world and use a variety of semiotic resources (modes producing signs).

## Methodology

Video recordings and partial multimodal transcriptions of one pair of students and a teacher involved in the research-action group were made and written documents distributed by the teacher were collected. The transcriptions (cf. Atkinson & Heritage, 1984) were done according to the following conventions (cf. Table 1).

Table 1. Conventions for multimodal transcriptions

| |
|---|
| Multimodal transcription conventions |
| ^ : rising tone |
| ‘ : falling tone |
| [ : overlapping speech |
| a : underlining words implies insistence |
| : ou ::: : a sound is drawn out |
| = : immediate chaining with next utterance |
| (.) : micropause |
| (3 s) : pause in seconds |
| (...) : a cut in the recording |
| (inaudible) : inaudible passage |
| ** gestures are described and shown in relation to the discourse that was spoken when they were performed |

Verbal and non-verbal behavior of the teacher and students were analyzed in relation to the classroom production. Written teaching-learning materials were used to define what type of resources students had at their disposal to build semiotic sets and bundles.

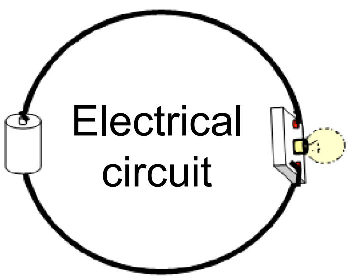
## Analyses and results

We present three short video extracts and their corresponding analyses. Before the third extract, we describe three groups of students' drawings. The first extract corresponds to the presentation of the analogy: the teacher defines each term and correspondence. In the second, the teacher constructs an experiment (lighting the bulb by connecting two wires to a battery) that corresponds to a physical object, a drawing of that object and a drawing of the experiment. In the third, the students attempt this same experiment.

### A teacher's complex yet coherent semiotic bundle

The first extract (second lesson) concerns an integration of two semiotic sets into a coherent semiotic bundle. The first semiotic set contains the model of the electrical circuit (Figure 1, left) and the second contains the analogy of the electrical circuit, built around a bakery that delivers bread to supermarkets (Figure 1, right).

Model of the electrical circuit




Analogy built around a baker delivering bread

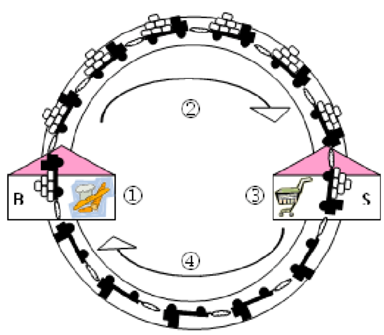



Figure 1. Two semiotic sets: the electrical circuit and the bakery analogy.

The elements of the analogy are presented below; each numbered element corresponds to the circled number on the figure at the right.

1. Each bakery always loads the same number of loaves of bread at into each delivery truck.
2. All delivery trucks move at the same speed. The speed of the delivery trucks is adapted to the demand of each supermarket.
3. The delivery trucks arrive at the supermarket where the loaves of bread are delivered, transformed and then sold to clients. All of the delivery trucks delivery all of their loaves of bread.
4. After delivery, each truck returns empty to the bakery in order to get a new load of loaves of bread.

The relations allowing the coherence of the semiotic bundle are provided by teacher's talk and gestures concerning each semiotic set and by written documents distributed at the beginning of the lesson (cf. Table 2 where T stands for teacher and S for student).

Table 2. Teacher discourse and gestures during her presentation of the bakery analogy.

| Time | Discourse | Gestures |
|---|---|---|
| 7'33 | T: the electric current can can circulate | traces a circle with her hand, then traces another |
| | Ok and this electric current what is it in the | rapidly traces a smaller circle with her hand |
| | delivery truck analogy^ | raises her right hand in order to indicate the change from the domain of electricity to the analogy |
| | S: the delivery [trucks | |
| | T: [why yes ^ | points her finger at a student |
| | S: the lineup of delivery trucks= | |
| 7'43 | T: =the lineup of delivery trucks very good | traces a circle with her hand |
| | it's the lineup of delivery trucks | traces a circle with her hand, then traces another circle |
| | uhh it's the fact that | traces a circle with her hand |
| | it's the circulation of the delivery trucks | traces a half circle, pauses at the bottom |
| | the trucks | finishes her circle |
| | are moving | traces a half circle with her hand, pauses between the bottom and the top, finishes her circle |
| | so the electric current circulates | traces a circle pauses between the bottom and the top |
| | ok it's all of the trucks that circulate very | traces a circle that is almost complete with her hand |
| 7'52 | good | |

Our videotape shows that gestures are identical when the reference domain changes (from electrical circuit to bakery analogy). The teacher makes a circular gesture to mimic both circulation of electricity (first semiotic set) and traffic of delivery trucks (second semiotic set). This circular gesture is repeated eleven times (cf. Figure 2), in association with the verbalization "circulation of delivery trucks" or "circulation of current". The rhythmic gesturing emphasizes the continuous aspect of the electrical circuit (delivery trucks) and allows the teacher to insist upon the systemic point of view. Moreover, she makes pauses corresponding to the positions of the supermarket (bulb) and of the bakery (battery). We argue that these similar gestures within two separate semiotic sets help students to integrate the analogues in a single and coherent semiotic bundle.

Figure 2. Teacher gesturing in order to make the link between circulation ofdelivery trucks and circulation of current.

However, problems begin for students at a much earlier stage. This paper's aim is to show that building complex semiotic bundles such as this one can be difficult for students unless each semiotic set they add to their bundle under construction (indeed each semiotic resource that they add to their set) is done in a coherent way. In order to illustrate this, we look at the bundle at an earlier stage, before the analogy has been introduced, when the objective is to build a simple circuit. In this context, a first semiotic set will consist of a drawing, written comments on that drawing, gestures about that drawing and a verbal description of that drawing, all done by the students. A second semiotic set will consist of handling the experimental apparatus, talking about handling it, performing gestures to show objects, all mostly done by the students alone. Having said that, semiotic sets and bundles can be co-constructed between teachers and students, adding to the complexity of our analytical viewpoint.

## A teacher's initial coherent semiotic bundle

Figure 3 shows an example of how an expert teacher is able to construct a coherent semiotic bundle from a variety of physical objects and functional representations of those objects (following her own instructions given to the students during the first lesson, as stated below):

1. Do a drawing that represents what happens inside the flashlight when it is turned on and is shining (a photo of this object is given – it's a familiar object for all students).
2. Draw diverse experiments with the proposed objects that will permit you to verify what happens inside the flashlight when it is shining.
3. After the teacher gives permission, ask for the necessary objects and perform the experiments.
4. Is your drawing for question 1 correct? If it isn't, do another one that shows what is happening inside the flashlight when it is shining.

Figure 3 shows that it is easy to link each part of the drawings with the experimental apparatus, because aspects and locations of objects (battery, bulb) are similar. These relations are also facilitated because drawing and experimental apparatus are built on the basis of the electrical diagram the teacher has in mind (far right of Figure 3). The teacher is thus able to move from one mode to another while maintaining coherency and building meaning. We also notice that the teacher does *not* draw the clamps of the crocodile clips in her drawing of the experiment, as they are not relevant to carrying out the experiment. As an expert, she is able to model the physical objects in a way that is pertinent for the experiment she will be carrying out, only paying attention to the characteristics that are relevant for that objective (see discussion on affordances in the next section).

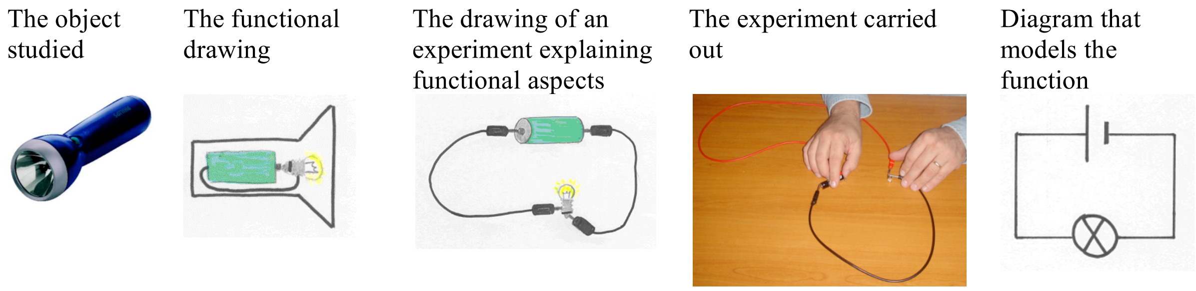


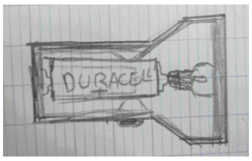

Figure 3. An *a priori* didactical analysis of multimodal and polysemiotic coherence for the simple circuit

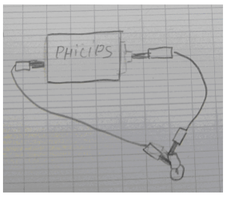
## Students' difficulties in constructing the initial semiotic bundle

We first show an analysis of three different student groups' drawings in order to illustrate the range of difficulties students have when they propose experiments that fit to their representation of what happens inside the flashlight when it is turned on and is shining (cf. Figure 4). One of the first difficulties students experience while building the semiotic bundle is connecting the experimental apparatus to the initial 'thought objects' that represent the flashlight parts in their drawing. From an expert point of view, the real object (the flashlight), the drawings and the experimental apparatus as well as discourse and gestures about these, form one semiotic set. On the other hand, we see that students already have difficulty relating these elements to each other, *within the semiotic set itself*, so adding another set to this one (e.g. the bakery analogy) to make a semiotic bundle seems, at this stage, to be out of these students' reach.

| Group# | The functional drawing | The drawing of an experiment explaining functional aspects | Comments |
|---|---|---|---|

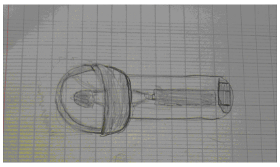
1

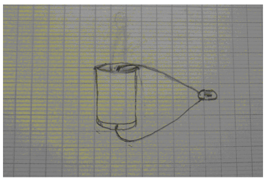



There is no wire in the functional drawing. Names of batteries are different from one drawing to another although the battery is the same. However, the drawing of the experiment is correct.

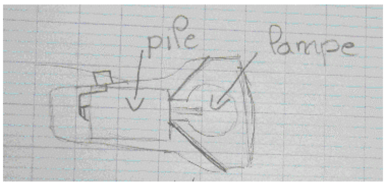
2

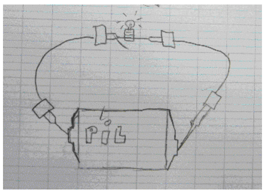




Students use one wire to connect either the battery to the bulb or the battery to itself. They tell their teacher that there is a button they can press to initiate battery functioning.

3




The functional drawing reproduces the general look of the bulb.
The drawing of the experiment takes into account the objects students have to use and the physical affordances of their experimental apparatus are evident in their drawing, even though this is not pertinent.

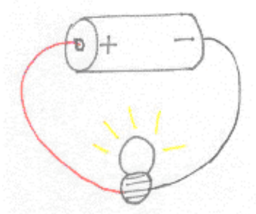
Figure 4. Example of students' polysemiotic incoherencies for the simple circuit.

The second extract (cf. Figure 5) concerns two students (Pierre and Jacques) who experience many difficulties in attempting to perform the simple circuit experiment as described above. Firstly, the students' drawing of the experiment is not correct. In order for the bulb to light up, one wire must be touching the very end of the bulb and the other wire must be touching the middle part (cf. the teacher's drawing in Figure 3).

So in this case, Pierre and Jacques are attempting to create coherency within an underlying meaning structure between two modes while trying to respect a drawing that will not help them. Secondly, the physical affordances of the crocodile clips (not shown in their drawing, but visible in the photo), entice the students to clamp them onto something, but the bulb does not allow clamping. Instead, one must just maintain the clamps stable (without trying to open them) so that they are touching the correct parts of the bulb and battery (cf. the teacher's experiment, above). According to Suthers, et al. (forthcoming), a given medium offers particular affordances (potentials for action in relation to the actor, following Gibson, 1977) of which salient affordances are expected to be the most pertinent (Norman, 1999). In this case, the crocodile clips were salient, but their function was not pertinent. Contrary to instructions, Pierre and Jacques do not redraw the experiment in an attempt to focus on how the electrical connections must be made, although this could have helped them.

The object studied

The drawing of an experiment explaining functional aspects




The experiment carried out a first time

Figure 5. Analysis of two students' multimodal and polysemiotic *in*coherence for the simple circuit.

6

Later on, another student shows Pierre and Jacques how to light the bulb with one wire. On this basis, they make several trials and suddenly, it works. Pierre says "it's magic!", but since he hadn't drawn any new representation of the experiment, this event remains at the level of the magic trick for him, because it is not integrated into the semiotic bundle. This interpretation is also illustrated by a final trial made with the teacher that does not work either. Figure 6 and Table 3 with the associated discourse of the teacher shows that Pierre has not yet understood how to hook the wires up to the battery. In fact, Pierre and Jacques followed their drawing, but this did not result in a lighted bulb and so they were stuck.

Table 3. Teacher discourse when a student helps him to hold the experiment.

| Time | Discourse | Gestures |
|---|---|---|
| 32'44 | T: no, not really here well we'll see that later on but it should not be placed on the same area of the bulb, | points to the wire on the bulb |
| | it should be on this area and on the contact | points to the lower side of the bulb and the bottom of the bulb |

Figure 6. Pierre is called upon to help the teacher hold the experiment.

## Conclusions and further work

In this article, we have used the notions of multimodal reformulation and the semiotic bundle to describe how meaning-making occurs when students and teachers speak, gesture, draw and manipulate objects. Choosing to look at meaning-making through the construction of such semiotic bundles allows us to meet two goals. Firstly, we render explicit teachers' expertise (i.e. their ability to seamlessly change modes, to select the appropriate characteristics of experimental apparatus to model in relation to a known objective and finally to coherently construct a complex semiotic bundle). Secondly we pinpoint where difficulties appear for students. In particular, if students erroneously construct a particular semiotic resource (e.g. the drawing for the experiment), they won't be able to build an underlying structure of meaning between it and the next semiotic resource they reformulate into (e.g. experimental apparatus). Secondly, the physical affordances of experimental materials (e.g. the crocodile clips) can hinder reformulation between a semiotic resource in one set and a semiotic resource in another (e.g. between drawing and experimental manipulation). Finally, although we did not show the data in this paper because of lack of space, some subsequent student interventions indicate that the multimodal reformulations produced by the teacher in order to broaden the semiotic bundle facilitate the understanding of the electric circuit: students reproduce similar gestures to those of their teacher and they discuss new metaphors based on the systemic point of view. We can conclude that it is useful to consider teaching sequences from the viewpoint of the *complexity* of the semiotic bundle that is being built. Each new semiotic resource and set that is added to the bundle must be constructed without error so that students are able to reformulate a given resource into another mode. Physical objects to be manipulated must be carefully chosen so as to not mislead students by material affordances that are not relevant to the task at hand. In sum, each step of the construction of each semiotic set that makes up a semiotic bundle can be seen as a checkpoint for student understanding. Future work will include the documentation of additional examples of representational mode changing and semiotic bundle building, both by teachers and students. We will also further explore how students can appropriate teachers' expert multimodal reformulation practices and the extent to which this affects their learning.

## References

Apothéloz, D. (2001). Les formulations collaboratives du texte dans une rédaction conversationnelle : modes d'expansion syntaxique, techniques métalangagières, grandeurs discursives manipulées, etc. In M.-M.

de Gaulmyn, R. Bouchard et A. Rabatel (dir.). *Le processus rédactionnel. Ecrire à plusieurs voix* (pp. 49-66). Paris : L'Harmattan.

Arzarello, F. (2004). Semiosis as a Multimodal Process. *Relime*, Numero Especial, 267-299.

Atkinson J.M. & Heritage J. (1984). *Structures of Social Action*, Cambridge, CUP.

Bécu-Robinault, K. (2002). Modelling activities of students during a traditional labwork. In H. Niedderer et D. Psillos (dir.), *Teaching and learning in the science laboratory* (pp. 51-64). Dordrecht: Kluwer Academic Publisher.

Bécu-Robinault, K. (2007). Modélisation et investigation autour d'une séquence d'électrocinétique : introduction d'une analogie pour expliquer. In L. Morge et J.-M.Boilevin (dir.). *Séquences d'investigation en physique-chimie* (pp. 117-129). CRDP Auvergne : Scéren,.

Brassac, C., Fixmer, P., Mondada, L., Vinck, D. (2008). Interweaving objects, gestures and talk in context. *Mind, Culture, and Activity*, 15(3), 208-223.

Buty C., Tiberghien A., Le Maréchal J.-F. (2004). Learning hypotheses and an associated tool to design and to analyse teaching-learning sequences. In M. Méheut et D. Psillos, (dir.). *International Journal of Science Education*, 26-5, 579-604.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2, 141-178. NJ: Erlbaum.

Chouinard, M.M. & Clark, E.V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30, 637-669.

Cosnier, J. (2000). V comme la Voix, les gestes, le corps. Dans (ed. Bernard Cerquiglini, et al.) *Tu parles!? : le français dans tous ses états*. Illustrations originales de Lewis Trondheim Publication. Paris : Flammarion p. 325-338.

Dicks, B., Soyinka, B. & Coffey, A. (2006). Multimodal ethnography. *Qualitative Research*, 6, 77-96.

De Gaulmyn, M.-M. (1987). Actes de reformulation et processus de reformulation in *La dame de Caluire*, pp. 83-98 Berne : Peter Lang.

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford, England: Pergamon.

Duit, R. & Von Rhöneck, C. (1998). Learning and understanding key concepts of electricity, in *Connecting research in physics education with teacher education* (A. Tiberghien, E. L. Jossem, J. Barojas Eds), *International Commission on Physics Education*. Published electronically at http://www.physics.ohio-state.edu/~jossem/ICPE/C2.html.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, and Knowing*. Hilsdale, NJ: Lawrence Erlbaum Associates.

Goldin-Meadow, S., Wagner Cook, S. & Mitchell, Z.A. (2009). Gesturing Gives Children New Ideas About Math**.** *Psychological Science.*

Kendon, A. (2004). Gesture : *Visible action as utterance*. Cambridge : Cambridge University Press.

Kress, G. and Van Leeuwen, T. (2001). *Multi-modal Discourse*. London : Arnold.

Lund, K. (2007). The importance of gaze and gesture in interactive multimodal explanation. *International Journal of Language Resources and Evaluation*. 41-3/4, 289-303.

Lund, K. & Bécu-Robinault, K. (forthcoming). La reformulation multimodale et polysémiotique comme aide à la compréhension de la physique. In A. Rabatel (dir.). *Analyse sémiotique et didactique des reformulations*. Presses universitaires de Franche-Comté.

McNeill, D. (1992) *Hand and mind: What gestures reveal about thought*. Chicago: Chicago University Press.

Norman, D. A. (1999, May-June). Affordance, Conventions, and Design. *ACM Interactions*, 6, 38-42.

Pineda, L. & Garza, G. (2000). A Model for Multimodal Reference Resolution. *Computational Linguistics*, 26 (2), 139 – 193.

Psillos, D., Koumaras, P. & Tiberghien, A., 1988, « Voltage presented as a primary concept in an introductory teaching on D.C. circuits ». *International Journal of Science Education. Vol.10 (1)*, p.29-43.

Roth, W.-M. & Pozzer-Ardenghi, L. (2005). Tracking Situated, Distributed, and Embodied Communication in Real Time. In M. A. Vanshevsky (Ed.), Focus on Cognitive Psychology Research: Nova Science Publishers, Inc.

Scott, P., Leach, J., Hind, A. & Lewis, J., (2006). Designing research evidence-informed teaching sequences. In R. Millar, J. Leach, J. Osborne et M. Ratcliffe (dir.). *Improving Subject Teaching: Lessons from Research in Science Education*. London: Routledge Falmer.

Suthers, D. & Medina, R. (forthcoming). Tracing Interaction in Distributed Collaborative Learning. In Puntambekar, Erkens and Hmelo-Silver (Eds.) *Analyzing Collaborative Interactions in CSCL: Methods, Approaches and Issues*. New York: Springer.

## Acknowledgments
We would like to thank the students and teachers who were filmed for this study and the educational administrators who accepted that such research be carried out in their school.

# Digital art-making as a representational process

Erica Halverson, UW-Madison, 1086 Educational Sciences, 1025 W. Johnson, Madison, WI 53706, erhalverson@education.wisc.edu

**Abstract:** In this paper, I take a distributed cognition framework to analyze the tasks of artistic production in terms of the external representations produced. Using data collected through case studies with four youth media arts organizations (YMAOs), I ask how the digital art-making process can be understood in terms of the construction of external representations and what function these representations serve in art-making. I analyze data on the process of making digital art in terms of the macro and micro tasks performed (Spillane, Halverson, & Diamond, 2001) in order to highlight external representation construction across organizations. Then I describe one micro task in depth to unpack the role the representations created serve in sensemaking. I find that digital art-making as structured in YMAOs seems to provide a robust, authentic environment for the construction of process-based external representations which has implications for content domains that value representation as a learning goal.

## Introduction

> I think that's what artists do. They struggle with their pieces, with what their pieces are about. They struggle with who they are and how they're going to show themselves (Street Level participant, 09.08).

Producing external representations that demonstrate mastery of a topic is often the desired outcome of participation in a constructivist-oriented curriculum. Performance-based assessment is predicated on the idea that the products of a learning process accurately represent what learners know and can do. While curriculum and assessment designers often struggle to construct tasks that afford learners the opportunity to engage in authentic production, art-making is fundamentally a representational task (Eisner, 2002; Langer, 1953). In producing art, we always strive to represent an idea, a feeling, a story to an audience using the tools of artistic media. In this paper, I describe how structured artistic production experiences provide adolescents with the opportunity to engage in a series of complex representational tasks. These representational tasks are consistent with a broad framework demonstrating the importance of external representations in sensemaking (Cox, 1999; Reisberg, 1987) as well as our understanding of the role of representation in creativity (Hasirici & Demirkan, 2007) and the artistic production process (Eisner, 2002).

The use of external representations for sensemaking in the context of schooling has thus far been situated in research on mathematics (e.g. Zhang & Norman, 1994) and science (e.g. Puntambekar & Goldstein, 2007). Philosophers of art (Langer, 1953) and scholars who describe the arts and cognition (Gardner, 1973; Eisner, 2002) take engagement with complex external representations to be a core part of artistic production. Despite a common emphasis on creating external representations as a method for students to demonstrate mastery, there is virtually no connection across research on artistic production activities and core subject areas. One reason for this may be the lack of common language for describing the value of representation for sensemaking in classroom-based academic tasks and in artistic production. The ability to evaluate and construct representations in physics, for example, is considered a high-level skill, and one that is essential for students to master. The arts, on the other hand, are often considered "extra" activities that, despite their direct parallels with scientific investigation (Gardner, 1973), do not warrant the same prestige. Understanding the role external representations play in artistic production processes can demonstrate the ways learning art and learning science are alike, breaking down the artificial separation between core and "extra" academic subject areas.

In order to understand the role of external representations in artistic production, I take a distributed cognition framework to analyze the tasks of the art-making process in terms of the external representations produced. Using data collected through case studies with four youth media arts organizations (YMAOs), I ask how the digital art-making process can be understood in terms of the construction of external representations and what function these representations serve in art-making. I analyze data on the process of making digital art in terms of the macro and micro tasks performed (Spillane, Halverson, & Diamond, 2001) in order to highlight external representation construction and use across organizations. Then I describe one micro task in depth to unpack the role the representations created serve in sensemaking. I conclude with reflections on the important role digital art production can serve in student learning and development with implications for content domains that value representation as a learning goal.

## The Role of External Representation in Sensemaking

Understanding the role of external representation in sensemaking begins with a distributed cognition perspective on thinking and learning (Hutchins, 1995; Salomon, 1993). Distributed cognition shifts the "locus of knowledge" from inside the individual to within and among actor-tool-activity networks (Salomon, 1993).

Understanding how people interpret and construct external representations using available tools over the course of systemic activity is a core task of research on distributed cognition (Cox, 1999; Zhang & Norman, 1994). Hutchins (1995) describes this research as, "the examination of the role of material media in which representations are embodied and in the physical properties that propagate representations across media" (p. 266).

In a distributed cognitive system, external representations function both as tools to think with (Norman, 1993; Zhang & Norman, 1994) and as tools for expressing complex ideas (Norman, 1983). Cox (1999) distinguishes between external representations that are provided for learners to assist in their understanding (presented or pre-fabricated) and the representations that learners construct to demonstrate their own understanding or to assist their own learning (self-constructed). While much of the research on the role of external representations in distributed cognitive systems has focused on learners' use of presented representations, Cox (1999) argues that constructing representations are of vital importance to understanding learning: "This activity represents more than a simple translation, for as Vygotsky observed, when signs (language, diagrams, etc.) are included in an action, they do more than facilitate manoeuvres that are impossible in the absence of the sign system. They fundamentally transform action" (p. 347). Reisberg (1987) described external representations as both *aides memoires* – expressions of understanding that extend human cognition – and *aides pensées* – tools with which to develop new understanding. In both cases, "one can learn by turning one's [internal] representations into stimuli – i.e. by externalizing them" (p. 288). How learners construct external representations of complex ideas and the value of this activity in building robust understanding is of great interest to educators and researchers who value a constructivist approach to learning.

### External Representation and the Arts

Artistic production provides a rich context for understanding distributed cognition and the role of external representation (Eisner, 2002; Gardner, 1973). Eisner (2002) describes artistic production as, "[the] transformation of a material to a medium. Materials *become* media when they mediate…[and] to convert a material into a medium is an achievement" (p. 80). Like any representational domain such as mathematics or writing, "getting smart" in the context of producing art, "means coming to know the potential of the materials in relation to the aims of a project or problem" (Eisner, 2002, p. 72). Art-making is fundamentally a creative process, and creativity is intimately tied to the production of external representations (Csikszentmihalyi, 1996). Hasirci and Demirkan (2007) describe mental images and external representations as "essential in investigations of creativity" (p. 262). Hayes (1989) describes the capacity to build the right representation in the right situation as a marker of creative expertise. There are strong parallels between the role external representations play in creative acts in art and in science. In both domains, creativity requires convergent and divergent thinking; the ability to construct representations that are both unique and recognizable is the marker of an accomplished artist and scientist (Gardner, 1973).

Artistic production is about the creation of representations. Creating art mindfully, that is learning how to construct and critically evaluate these representations, requires scaffolded instruction. In youth arts organizations, participants are engaged in authentic artistic production tasks, yet they are guided in these authentic tasks by explicit pedagogical practices. Eisner (2002) describes student artists in the context of formal instruction as, "work[ing] meaningfully on the creation of images" (p. xii). Research on artistic production in youth arts organizations has demonstrated that youth are highly invested in the meaningful construction of images with explicit attention to how these products will be received by a public audience (Heath, 2000; Soep, 2006). In my research, I am interested in how youth learn to engage with and create external representations over time. I follow this process in youth arts organizations because they offer youth explicit instruction (much like schools) toward the goal of the public sharing of art (much like professional artists). It is in this context of these organizations that I take up questions of external representations and artistic production.

## Research Methods

Broadly, I am interested in the role self-constructed external representations play in learning, particularly in the arts. While seminal research on arts-based learning has broadly considered the importance of representations in art-making (Eisner, 2002; Gardner, 1989), this research has not taken an explicitly distributed cognition perspective on the art-making process in order to understand how representations are constructed and the functions they serve. To initiate this research trajectory, I posed the following research questions: (a) What are the representational tasks involved in producing autobiographical digital art? (b) What function do those representational tasks serve emerging art-makers in youth media arts organizations as they create products to share with a public audience?

### Data Collection

To address these questions, my research team conducted four instrumental case studies (Stake, 2000) with YMAOs across the United States. Youth arts organizations work in a variety of media including live theatre, radio, digital story, film, and spoken word poetry. While all of these media afford youth the opportunity to engage in the representation of self through art, we chose YMAOs that work in the digital media arts, with a focus on film. This allowed us to hold the representational medium constant as we explored how youth constructed and used external representations. Table One briefly describes the four organizations.

Table 1: Four case study organizations.

| | Location | Production cycle length | Type of program | Time of data collection | # of youth participants |
|---|---|---|---|---|---|
| Appalachian Media Institute | Whitesburg, KY | 3 months | Summer internship | Summer 2008 | 12 |
| In Progress | St. Paul, MN (headquarters) | Varied | Ongoing | Summer 2007 | 10 |
| Reel Works Teen Filmmaking | New York, NY | 6 months | Semester | Fall 2007 – Spring 2008 | 12 |
| Street Level Youth Media | Chicago, IL | 3 months | Summer internship | Summer 2008 | 12 |

At each organization, we traced one production cycle – from participants' initial entry into the organization to the final presentation of their work. In identifying a production cycle as the frame for our cases, we created further consistency in our data collection not by the amount of time we spent with each organization but by the organizational structure provided for participants to produce their digital art (Table 1). In order to capture the production cycle at every organization we collected a variety of qualitative data:

- *Documentation of the process in action*. Most of this documentation was in the form of ethnographic observation, though we also obtained video documentation at various points across the production cycle.
- *Artifact collection*. We collected all artifacts youth created around the digital production process including application essays, journals, group brainstorming sheets, worksheets, edited/unedited video footage, and blog entries. We also collected curricular materials used by organizational leaders and individual workshop facilitators.
- *Interviews* with participants, organizational leaders, facilitators, and mentors.

Data collection was iterative; we employed a constant comparative method across our case studies (Glaser & Strauss, 1967). We began with two open-ended case studies (*Reel Works* and *In Progress*) and then returned to the field a year later to conduct two more directed case studies (*AM*I and *Street Level*) based on our initial interpretations. In our open-ended case studies, we participated in as much of the production cycle as we could, and collected all artifacts that seemed relevant. While we did not initially conduct a full analysis of all representations produced, we did identify when and where opportunities for construction occurred to inform the design of our second round of case studies. In this second round, we directed our efforts towards capturing youths' representation construction – both the process of constructing and their use of these representations in the context of the production cycle. We also asked organizational leaders more pointed questions about the relationship between identity and representation in our second round of interviews.

## Data Analysis

In a distributed cognitive system, the way actors make sense of their work can be analyzed in terms of the tasks they must complete and the artifacts generated in the completion of these tasks. In this analysis, I employ Spillane et al.'s (2001) analytic method for understanding distributed cognitive tasks, breaking down the digital art-making process into macro and micro tasks. In my previous work, I have documented autobiographical artistic production as a "dramaturgical process" – the telling, adapting, and performing of narratives of personal experience (Author, 2007, in press). Using an open-ended coding process (Corbin & Strauss, 2008) of observational data, interviews and artifacts created during the production cycle across all four case studies, I coded all the representational tasks we captured as one of three macro tasks: (a) Telling stories and/or figuring out what story to tell; (b) Adapting stories into filmic representations or; (c) Presenting products to an outside audience.

Once I classified all data that concerned the building of an external representation into these three categories, I conducted a second round of grounded coding where I assigned the data to a "micro task" within one of these macro tasks. Since tasks demonstrate the distributed interrelationships among actors and artifacts in sensemaking (Spillane et al., 2001), I considered an activity a micro task if participants had to create a unique external representation as an integral part of the activity. Having identified these micro tasks, I explored them further in analysis by identifying relevant properties and dimensions (Corbin & Strauss, 2008) of these micro tasks across organizations and by building theoretical conceptual matrices that highlighted analytic points of interest across the data.

## Findings: Representational micro tasks in digital art-making

Given the central role artifacts play in analyzing distributed cognitive tasks, my analysis affords opportunities to see these cognitive processes as representational tasks and to understand the tasks at a level of depth previously unexplored in cognitive analyses of artistic production. In this section, I first describe the representational tasks youth engage in as they created digital art at the four YMAOs. Table 2 provides a summary of all the micro tasks and their prevalence across organizations. In the narrative sections below, I reference the names of tasks by an individual data citation (field note, interview, journal entry, worksheet) though multiple forms of data demonstrated the prevalence of these tasks and their relevant artifacts. I then describe one representational task in depth: the pitch. In doing so, I analyze the role this task and accompanying external representations serve in sensemaking.

Table 2: Micro tasks in the process of digital art production.

| Type of task | Format(s) | AMI | IP | RW | SL |
|---|---|---|---|---|---|
| *Macro task: Telling stories* | | | | | |
| Application essay | Written text | √ | | √ | √ |
| Application interview | Oral conversation | | √ | √ | |
| Warm-up ideas | Written text; Multimodal | √ | √ | √ | √ |
| Pitch | Oral presentation<br>PowerPoint or other supplementary text | √ | √ | √ | √ |
| First treatment/shooting script/shot list | Written text | √ | √ | √ | √ |
| *Macro task: Adapting stories* | | | | | |
| Prose script | Written text | √ | | √ | |
| Rough cut(s) | Multimodal | √ | √ | √ | √ |
| *Macro task: Presenting final products* | | | | | |
| Final video | Multimodal | √ | √ | √ | √ |
| Artist statement | Written text | | | | √ |
| Public showing | Multimodal; Oral conversation; Written text | √ | √ | √ | √ |

### Telling Stories

In the context of producing digital videos, "telling stories" as a macro task is accomplished when filmmakers are ready to begin shooting footage for their films. Across all four organizations, there were five different micro tasks that youth participated in.

*The application essay*. In three of the four organizations youth were required to submit a written application. Applicants were given a series of guiding questions to respond to, including questions about what stories they might like to tell with film, as well as reasons for wanting to join the program.

*The application interview*. Two of the four organizations required youth to participate in one-on-one interviews with the program's directors prior to beginning the process. At Reel Works, youth had also written applications so the interviews were supplementary. At In Progress, interviews were the director's opportunity to talk with youth prior to their beginning the process about their goals and film ideas.

*Warm-up ideas*. While some youth came to the organizations having already decided what film they want to make, many had not. Youth produced a variety of artifacts that were "warm-ups" for their final product including short essays in response to questions such as, "Why do you like art and making stories? What are you going to do this year" (IP FN 07.31.07) or writing prompts such as, "I really miss…" and "My greatest passion in life [is]…" (RW Journals Fall 2008). Participants also produced more focused representations based on guided worksheets. Alternately, youth produced mini-films in a short production cycle such as "community-based interviews" (AMI FN 06.08) or "past assignments" (SL FN 07.01.08). The resulting artifacts were multi-modal, video-based projects that were accomplished in a short period of time and are given feedback by the adult facilitators.

*Pitch.* Across organizations youth were required to "pitch" their ideas to a critical group, either within their community or to some outside group. Youth produced a wide variety of artifacts during the pitch, depending on the nature of the task as structured by the organization. The pitch is described in greater depth below.

*First treatment.* Before participants began shooting their films, they created a representation of their plan – referred to as a "first treatment," (RW FN 07.20.07) "shot list," (SL Journals 07.08) "essays," (IP FN

08.01.07) and “proposals” that include who they will interview, shot list, b-roll footage, why they want to make the film and its importance (AMI Interview 04.08).

## Adapting Stories

To distinguish between the “telling stories” macro task and the “adapting stories” macro task, I divided the work by representations that were created before and after footage for the final product was shot. In some cases youth used footage they created in their “warm-ups” for their final product, blurring the line between the two macro tasks. Additionally, films that were heavily interview-based had a clearer distinction between pre-/post-shooting artifacts since none of the dialogue was crafted in advance. For films with narrative elements, some of the adaptation work was accomplished in advance of shooting. Despite this messiness, there were two micro tasks unique to the adaptation function.

*Prose script.* Participants at *Reel Works* generated scripts for their films after they shot the majority of their footage, giving them a roadmap for how to edit their films and a broad sense of how the final product would look. *AMI* had their participants create “paper edits…where you write down what you think your video will look like and what each part’s going to be like. Like some people do it minute-by-minute or just in minute chunks…like what scenes are going to look like and what each character’s going to bring” (AMI Interview 07.16.08).

*Rough cut.* Participants produced one or more “rough cuts” before they presented a final product. Typically, these rough cuts were used in critique sessions with peers and/or mentors. Rough cuts are multimodal since they represent the film to that point – they may be missing certain key modes (such as soundtrack) or they may contain all modes in basic form.

## Presenting Products

A core component of any performance-arts production process is the final presentation to an outside audience. Since audience is a critical component of artistic production, all of the micro tasks in the presentation function focused on creating a sharable product that can be communicated to outsiders. Three unique representational tasks emerged.

*Final video.* All participants produced a video, either alone or with a group. The final video is the ultimate external representation – the permanent instantiation of how an individual or group of youth chose to share their story with an audience. In other work I describe how these final products can be analyzed as standalone multimodal representations (Halverson, in press).

*Artist statement*. At *Street Level* youth created “artist statements”, written text that accompanied their videos describing who they are and what their piece is about.

*Public showing.* In addition to completing a video as a standalone piece of work, youth participate in a variety of public events where they have an opportunity to engage in a discussion about their work with audience members, either in the form of a gallery opening (Street Level) or a screening with a post-show “talk back” (other three organizations).

## One Micro Task In-Depth: The Pitch

While the above description outlines the representational tasks involved in producing digital art, I am also interested in the role these representations serve in helping youth to understand the art-making process. In order to explore the role representations serve art-makers, I chose one task and its accompanying representations to analyze in further depth. Table 3 provides a more detailed breakdown of “the pitch” at each organization.

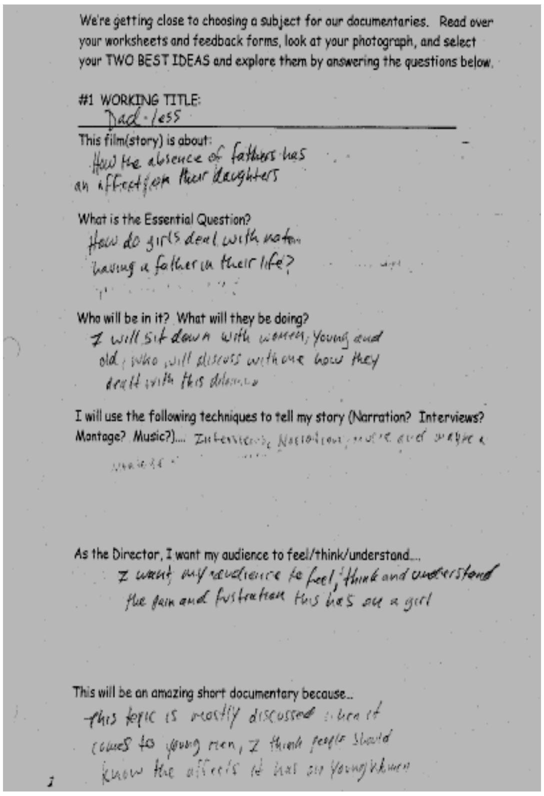
Table 3: Pitch micro task across organizations.

| | Micro-task | Artifacts produced | People involved |
|---|---|---|---|
| AMI | Brainstorming session | Ideas on butcher block paper<br>Oral presentation<br>List of ideas to vote on<br>Vote tally for project ideas | Participants<br>Instructors<br>Mentors |
| In Progress | One-on-one meeting | Journal notes | Participants<br>Director |
| Reel Works | Formal story pitch | Pitch worksheet<br>Oral presentation | Participants<br>Instructors<br>Mentors<br>External experts |
| Street Level | Group presentation | PowerPoint<br>Oral presentation | Participants<br>Instructors |

The artifacts generated for these pitch sessions provide opportunities for youth to engage with their film ideas by a) expressing textually and orally what story they want to tell and why and b) engaging with experts and peers through reflective critique. The differences in task and artifacts created are isomorphic representations that are implementations of the same abstract goal (Zhang & Norman, 1994).

The *Reel Works* pitch process is the most traditional example of "pitching" movie ideas to sell to outsiders. Youth participants traveled to Time-Warner studios in New York City to share their ideas with their peers, mentors, and directors, and a panel of professional producers. Youth were instructed to treat this meeting as if they were selling their movie ideas to producers, providing a context for their presentations. In addition, they were given a worksheet to guide their thinking about the development of their ideas (Figure 1). Some youth used the worksheet directly, filling in responses to worksheet questions and using these responses to give their pitches. Others used the questions on the worksheet as guidelines for the construction of a narrative description that they created in a journal. Youth used these artifacts to read from as their delivered their pitches, though they also answered questions from the audience extemporaneously.

The worksheet provided space to reflect the content of the story – e.g. "*This film(story) is about*: The life of my grandfather; Through good times and bad. How I handled the ups and downs." It also provided space to think about the representational medium of film – e.g. "In this film, I would be the narrator. I will also interview my parents and other artists who appreciate the art of photography like I do. There will definitely be music in the beginning of this film." Finally, participants reflected on the meaning of their film: "*As the Director, I want my audience to feel/think/understand*…where I'm from, try and put theirselves [sic] in my situation, be inspired at the end of my film. I want people to go out and change their situations, my film should be influencial [sic]." The questions posed on the worksheets encourage youth to think about the content of their film, their use of film as a medium, and why their film is important. This explicit attention to content, structure, and impact resulted in external representations, in the form of written statements on their ideas, how they will be carried out and the functions they may serve.


We're getting close to choosing a subject for our documentaries. Read over your worksheets and feedback forms, look at your photograph, and select your TWO BEST IDEAS and explore them by answering the questions below.

#1 WORKING TITLE:
Dad-less

This film(story) is about:
How the absence of fathers has an affect on their daughters

What is the Essential Question?
How do girls deal with not having a father in their life?

Who will be in it? What will they be doing?
I will sit down with women, young and old, who will discuss with me how they dealt with this dilemma

I will use the following techniques to tell my story (Narration? Interviews? Montage? Music?)....

As the Director, I want my audience to feel/think/understand....
I want my audience to feel, think and understand the pain and frustration this has on a girl

This will be an amazing short documentary because...
this topic is mostly discussed when it comes to young men, I think people should know the affects it has on young women

Figure 1. *Reel Works* pitch worksheet.

The oral presentation afforded youth the opportunity to engage in reflective critique that pushes these ideas further. These critique sessions focused on both practical issues of filmmaking and about the function of film as a medium in communicating a complex idea. For example, in pitching *Rupture,* the film about absentee fathers referenced in Figure 1, the filmmaker, instructor, mentor, and an external expert have an extended exchange about how to construct this piece so that it shows certain universal themes yet maintains her individual experience. As the instructor describes: "your story is one that many, many people have experienced but when you bring us through the story through your eyes, of walking through those doors, of what you were thinking and then of what your mother was thinking as you came through those doors" (RW pitch meeting, 03.08).

In contrast, at AMI*,* the focus of the pitch is for youth to sell their ideas to one another. Since films are made in groups, the pitch session serves as an opportunity to select three ideas they will choose to make. Youth engage in an extended brainstorm session where they generate their film ideas and try to draw connections across ideas (Figure 2). Every person has a chance to contribute their ideas, to describe the idea they contributed to the brainstorming session, and to indicate which of the ideas are their favorites. The butcher block paper is used to generate ideas and to draw connections across stories.

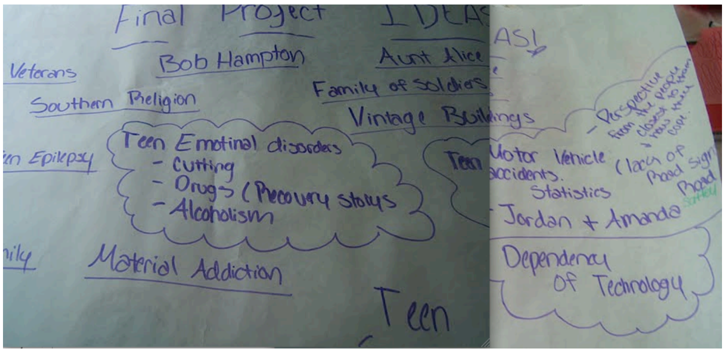



Figure 2. Final project brainstorm at *AMI.*

Once three ideas have been selected through a blind voting process, and youth are assigned to groups based on their own interests and mentors' construction of productive teams, these new groups devise a "proposal" for their film. The proposal consists of basic structural ideas, whom they will interview, a preliminary shot list, and a description of the film, as well as an assessment of who the audience for the film will be. Similar to Reel Works*,* AMI participants are given a series of questions to answer about their prospective film; they use "sample proposals" as models to construct their own (AMI FN, 06.27.08).

While the content of the pitches are constructed representations of participants' ideas, the structure for the representations is created by the organizations. By prompting youth with questions to answer, models to follow, and formats to fit ideas into, the organizations provide a structural scaffold within which participants can craft representations of their ideas. The pitch happens relatively early in the process, usually about one third of the way through the production cycle. At this point, youth are likely still struggling with core questions around why their ideas matter and whom these ideas may reach. Structuring the representational task in this way facilitates youth engagement with these questions and requires that they construct external representations, in the form of text and oral presentations that describe their emerging understanding of the value of their ideas and how these ideas can be communicated to outsiders. Across the pitch meetings, instructors and mentors used participants' pitches to help youth think in depth about the generalizability of their individual narrative ideas.

## Discussion: External representations in art-making and beyond

Analyzing the micro tasks embedded in the telling, adapting, and performing of narratives of personal experience demonstrates that learning to make art in the context of YMAOs is fundamentally a distributed, representational endeavor. YMAOs structure opportunities throughout the process for participants to create artifacts - written, oral, and multimodal – that demonstrate youths' evolving understanding of the story they want to tell and how the medium of film affords them the opportunity to represent this idea. Building these representations provides opportunities for sensemaking and reflection both in the representations themselves and in the discourse generated around these representations. Each microtask results in an external representation that affords the opportunity to explore the relationship between the idea and how the medium of film can express this core idea. The films themselves are multimodal representations of sophisticated ideas (Halverson, in press); seeing the microtasks from initial conception to final product allows us to understand how the final representation comes to be and whether and how youth are mindful of what their artistic representation conveys.

This type of distributed, representational system is not limited to digital art-making. Artistic production is a representational act, and learning to create art in an organizational context allows emerging artists to understand how representation affords reflection on art and meaning. The representational tasks scaffold the art-making process for youth to become reflective artists also look a lot like the representational process we want our students of science and mathematics to engage in.

## Educational Significance

Understanding artistic production in terms of external representational tasks is significant to researchers, practitioners, and designers who are interested in constructivist, project-based learning environments, especially those that engage learners with 21st Century skills. A decade ago, Cox (1999) described the growing importance of representation construction in progressive learning environments: "As educational technology becomes more and more integrated into the curriculum and as the variety and sophistication of data visualization and external representation techniques proliferate, the issue of training students to use ERs effectively is likely to increase in importance" (p. 360). Digital art-making as structured in YMAOs seems to provide a robust, authentic environment for the construction of process-based external representations.

Constructing external representations of process over time is a crucial component of YMAOs as designed environments. Despite this, research on the design of learning spaces that promote the development of complex external representations has been left to the sciences particularly physics, engineering, and the biological sciences. My analysis confirms Eisner's assertion that we need to allow artistic production to enter the conversation around external representations, complex cognition and the design of learning environments: "the tasks that the arts put forward – such as noticing subtleties among qualitative relationships, conceiving of imaginative possibilities, interpreting metaphorical meanings the work displays, exploiting unanticipated opportunities in the course of one's work - require complex cognitive modes of thought" (Eisner, 2002, p. 35).

Finally, narrative media such as digital art-making may be especially potent for adolescents and emerging adults who are involved in a psychosocial identity development process. Identity development has been described as the negotiation of a relationship between cultural and personal narratives (Hammack, 2008) – explicit engagement with these narratives through the dramaturgical process could facilitate positive developmental outcomes for participants. Artistic production that focuses on autobiographical experiences opens up opportunities for the representational process to engage youth in identity exploration. Future work will focus on how constructing external representations over time is intricately connected to identity development.

## References


Corbin, J. M. & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Thousand Oaks, CA: Sage Publications.

Cox, R. (1999). Representation construction, externalized cognition and individual differences. *Learning and Instruction, 9*, 343-363.

Csikszentmihalyi, M. (1996). *Creativity.* New York: Harper Collins.

Eisner, E. (2002). *The arts and the creation of mind.* New Haven, CT: Yale University Press.

Gardner, H, (1973). *The arts and human development.* New York: Harper Collins.

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Halverson, E. R. (in press). Film as identity exploration: A multimodal analysis of youth-produced films. *Teachers College Record, 112*(9).

Hammack, P. (2008). Narrative and the cultural psychology of identity. *Personality and Social Psychology Review, 12*(3), 222-247.

Hasirci, D. & Demirkan, H. (2007). Understanding the effects of cognition in creative decision making: A creativity model for enhancing the design studio process. *Creativity Research Journal, 19*(2-3), 259-271.

Hayes, J. R. (1989). Cognitive processes in creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity.* New York: Plenum.

Heath, S. B. (2000). Seeing our way into learning. *Cambridge Journal of Education, 30*(1), 121-132.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

Langer, S. (1953). *Feeling and form.* New York: Charles Scribner's Sons.

Norman, D. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

Norman, D. (1993). *Things that make us smart.* New York: Double Day.

Puntambekar, S., & Goldstein, J. (2007). Effect of visual representation of the conceptual structure of the domain on science learning and navigation in a hypertext environment. *Journal of Educational Multimedia and Hypermedia, 16*(4), 429-459.

Reisberg, D. (1987). External representations and the advantages of externalizing one's thoughts. *The Ninth Annual Cognitive Science Society* (pp. 281-293). Hillsdale, NJ: Lawrence Erlbaum.

Salomon, G. (1993). *Distributed cognitions.* New York: Cambridge University Press.

Soep, E. (2006). Critique: Assessment and the production of learning. *Teachers College Record, 108*(4), 748-777.

Spillane, J., Halverson, R., & Diamond, J. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher, 30*(3), 23-28.

Stake, R. (2000). Case studies. In, N. K. Denzin & Y. S. Guba (Eds.), *Handbook of qualitative research*. Thousand Oaks, CA: Sage Publications.

Zhang, J. & Norman, D. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18,* 87-122.

# Kindergarten and First-Grade Students' Representational Practices While Creating Storyboards of Honeybees Collecting Nectar

Joshua A. Danish, David Phelps, Indiana University, 201 North Rose Ave, Bloomington IN, 47405
Email: jdanish@indiana.edu, davphelp@indiana.edu

**Abstract:** A productive approach to studying the role of representations in supporting students' learning of science content is to examine their representational practices. The current study examines kindergarten and first-grade students' representational practices in a similar context—the creation of storyboards—both before and after a curricular intervention in order to highlight those aspects of their practices that changed while engaging in a superficially similar task. Analysis of the students' storyboards reveals considerable improvement after the intervention. Analysis of the students' practices as they changed over time is also presented by examining the students' interactions, with a focus on their discussions of the science content and the representations themselves.

There is no doubt that students' ability to create representations such as drawings and graphs is central to their ability to learn science in school (DiSessa, 2004; DiSessa, Hammer, Sherin, & Kolpakowski, 1991; Lehrer & Schauble, 2006). One reason for the centrality of representations in science classrooms is that specific representational forms can help students more easily engage with and communicate complex ideas (Roth & McGinn, 1998). For this reason, representations are also instrumental in the work of professional scientists (Latour, 1988; Lynch, 1988), and thus many attempts to teach students how to engage in authentic science necessarily include helping them to create representations. After all, students do not always create representations that are immediately productive for learning and expressing science content. Rather, students need support in order to create productive and accurate representations regardless of whether they are being asked to create canonical representations (i.e., graphs) or invented representations (i.e., drawings or diagrams from observation). One method that has been used to examine the way that the students' context can support their creation of representations is to study students' *representational practices* as they engage in creating and refining representations in science class. The practice approach highlights the relationship between students' representational actions, their knowledge of the content, and the features of the context that enable and constrain those actions (Danish & Enyedy, 2007; Hall, 1996; Roth & McGinn, 1998).

Prior examinations of students' representational actions from a practice perspective have proven effective in several accounts of how representational practices contribute to or explain student learning within science and math (c.f., Cobb, Stephan, McClain, & Gravemeijer, 2001; Danish & Enyedy, 2007; Hall, 1996; Hall & Rubin, 1998; Roth & McGinn, 1998). However, there are two key gaps in the current literature that this paper aims to fill. First, examinations of students' representational practices typically focus on the way that these practices change in different contexts. This is, in part, to examine how students' movement through a range of representational activities supports them in learning new content. While this is a fruitful approach, it does not allow for a systematic comparison of students' practices as they change over time with respect to a single representational form. Second, this literature largely overlooks the representational practices of early elementary school children (for some exceptions see Danish & Enyedy, 2007; Lehrer & Schauble, 2000). Rather than neglecting the representational practices of young children, we argue that they are a crucial population to study because the practices that students develop early in their school careers will be important in supporting their later science and representational activities. Moreover, early elementary students spend a great deal of time learning by drawing, sculpting, and enacting their ideas in science class, and so it behooves us as a field to better understand these processes in order to inform research and teaching with representations.

This paper aims to begin filling these gaps in the literature by documenting the representational practices of 42 kindergarten and first-grade students in service of learning how honeybees collect nectar. The students were asked to create storyboards at two points in time: before and after a curricular intervention where they learned about honeybee content and were also encouraged to request and provide representational feedback (for the larger study upon which this analysis draws see Danish, 2009a; Danish, 2009b). First, to ground our discussion of the students' representational choices, we present a brief analysis of their representations (the storyboards) to highlight the fact that these did in fact improve over the course of the study. Then, video of students' interactions were analyzed to document the students' representational practices by identifying what their discourse revealed regarding the issues that they saw relevant in their representational practices. Specifically, their talk was coded along multiple dimensions to identify the prevalence and circumstance of talk about the science content and the representations themselves.

## Representational Practices

Simply put, practices refer to the patterned way of acting that people develop over time (Lave & Wenger, 1991). More specifically, Hall and Rubin (1998) further define *representational* practices as "making, using, and reading conventional representational forms," (p. 228). Hall and Rubin also caution us that these representational practices are "less stable and more detailed than many textbooks or studies of instruction presume" (ibid). The present study responds to this concern by documenting the details of students' representational practices as well as examining the process through which these practices change.

## Examining Representational Practice.

One hallmark of the practice approach to studying students' actions is that it recognizes the importance of context in shaping those actions (c.f., Roth & McGinn, 1998). To this end, most accounts of students' practices tend to at least implicitly recognize that there are features of the practice that are relatively stable across moments, and features that are highly dependent upon the moment being observed.

*Stable aspects of students' representational practices.* There are two ways of explaining stability, or observable patterns in students' representational practices. The first is to recognize that there are local social patterns, or norms that are stable within a particular setting. The second is to assume that individual participants "carry" the necessary knowledge across moments. Cobb et al. (2001) refer to the social aspect of this equation as norms, and the individual aspect as a psychological correlate, and argue that the two are mutually co-constructed and cannot be separated. Simply put, each individual has certain beliefs and knowledge about the practices and the content to which they refer, which include an understanding of the social norms surrounding it. At the same time, the social norms have grown out of and account for individual knowledge of the local participants. By definition norms are non-deterministic; students may or may not selectively follow them though one can typically identify norms by the way that participants mark their violation (Cobb, et al., 2001; Danish & Enyedy, 2007; Hall & Rubin, 1998).

In related research, diSessa and his colleagues identified "meta-representational competencies" which refers to students' understandings that apply across multiple representational forms, such as a belief in the need for parsimony (DiSessa, 2004; DiSessa, et al., 1991). Furthermore, these competencies, they argue, develop over time as students engage in representing in school (and other) contexts (Sherin, 2000). While these two approaches to locating stability in students' representational practices—the one focused on the social norms, and the other upon individual competencies—come from different theoretical perspectives, the result is the same; they describe patterns in how students engage with representations and suggest that these patterns, which are learned over time, have important implications for what students do with the representations that they encounter. Thus, it is important to document the patterns in students' representational actions, regardless of whether or not they appear to originate in social norms or individual preferences and competencies.

*The role of context in representational practice*. Of course, students' representational practices are also defined by their context of use, which may be defined in terms of the nature of a specific representation, of the task, or even of the other people present (DiSessa, et al., 1991; Hall & Rubin, 1998; Roth & McGinn, 1998). For example, students may not assume that a representation needs to be self-explanatory when they know that their teacher will allow them additional time to explain it further (Danish & Enyedy, 2007). There is, however, a reciprocal relationship between practices and context; individuals are presumed to see the context differently as a result of their changing practices. In fact, many of the studies described above rely on this to examine how students developing practices, combined with new contexts, can lead to key content learning. Therefore, in order to document students' representational practices, we attempt to identify those aspects that are relatively stable, the representational context, and the interaction between the two.

## Documenting Representational Practices

Traditionally, examinations of representational practices have focused on how they change over time along with shifting instructional activities. Researchers then identify students' practices by examining talk (and actions) as the students engage in creating, debating, and modifying representations. Hypotheses are then developed regarding what types of behavior are normative. These are vetted against the corpus of data to look for seeming violations of those norms (for an in-depth description of one such approach see Cobb, et al., 2001). This approach makes sense, given that the goal of these prior studies was to understand how specific practices, when paired with specific instructional exercises, resulted in students learning about the content being studied. Unfortunately, a side-effect of this approach is that these studies provide considerably less guidance to researchers and teachers whose aim is to support more general representational activities across a range of content domains. Furthermore, while patterns in the practice are described, it is difficult to distinguish patterns from the context in which they are visible for the very reason that the researchers are shifting those contexts in order to pursue specific instructional goals. For these reasons the present study aims to examine students' representational practices in a relatively stable context—the creation of storyboards. Of course, as students' practices shift students may be expected to relate in different ways to a superficially similar context. However, by holding the task stable, we are able to gain a richer understanding of the inter-relationship between the students' changing practices and the context of their use. Finally, by shifting our focus away from the process through which students' representational practices support (or not) their learning of specific content topics, we are able to more thoroughly document the entire range of their practices.

### What do we know about young children's representational practices?

Well before entering kindergarten, students are able to produce representations and understand their symbolic functioning (DeLoache & Burns, 1994). Studies have also shown that young children are aware of the context of their representational choices, and can adjust their interpretation or creation of a representation in response to their audience and the context (Callaghan, 1999). In fact, it was even demonstrated that kindergarten and first grade students appeared to recognize and balance between multiple competing influences on their representational choices including individual preferences, local norms, and the affordances of the immediate context (Danish & Enyedy, 2007). Furthermore, it has been demonstrated that young students are able to develop a rich sense of the requirements of a model for scientific activity, and to adjust their models accordingly as they engage with learning new content (Lehrer & Schauble, 2000). In short, there is reason to believe that young children are capable of rich and elaborate representational practices, and that these practices influence the role that representations play in supporting young students in learning about science content.

## Methods

The data presented in this paper comes from part of a larger study known as the BeeSign study, a 10-week curriculum intervention described in detail elsewhere (Danish, 2009a, 2009b). However, the prior analyses examined student content learning based on pre- and post interviews and primarily focused on the classroom intervention. The present study extends this work by examining the storyboard activity administered before and after the larger study. The study as a whole, and the storyboard in particular, will be described below.

### Participants

This study took place with two mixed-age kindergarten and first-grade classrooms (most of the students were between 5 and 7 years old) in a progressive California elementary school. The school as a whole has a range of family incomes from $35,000 or less per year to $250,000 or more per year with a demographic composition of approximately 47% Caucasian, 27% Latino, 13% African American, and 13% Asian. There were 42 students who consented to participate in the study (22 boys, 20 girls).

### Procedures

The students in this study participated in the BeeSign curriculum as their science curriculum for over 10 weeks. There were typically two one-hour-long sessions per week, for a total of 19 sessions. Over the study, the students were taught about how honeybees collect nectar through a series of inquiry and representational activities. In the representational activities, the teachers encouraged the students to provide and solicit feedback from their peers.

The students began the study by engaging in a pre-test, followed by the storyboard activity to be described below. They then engaged in the curriculum activities, followed by a final storyboard activity and then a post-test interview. The interviews were scored in terms of students' understanding of honeybees in terms of the structure, behavior, and function of the system based on the work of Hmelo-Silver and colleagues (2004) who had documented that novices tended to focus on the superficial structures and behaviors of a system at the expense of the functions. The students in the BeeSign study showed statistically significant gains from the pre- to post-interview in terms of the structures, behaviors, and functions that they described (Danish, 2009a, 2009b).

#### The storyboard activity

The present analysis focuses on the storyboard activity that took place both before and after the curricular intervention. During this activity the students were divided by the teachers into small groups of 3 to 6 students who sat together at a table. One half of the groups remained stable from the pre- to post- test, and the other half did not because one of the two teachers chose to allow the students to form their own groups in the post-activity. At both time points the students were asked to draw a storyboard about "how bees get food" and were shown how to use a storyboard to represent a series of steps in a process. The students were given a template that included 4 boxes in which to draw the storyboard, with several lines beneath each box in which they could write descriptions of the images. The students were allowed to talk and solicit feedback from their peers. However, the teachers did not actively request this (as they had during the curriculum intervention) in order to capture how frequently the students did this of their own accord. There were 6 groups videotaped at each time point, but only 10 of the groups (5 in the pre condition, and 5 in the post) had audio and video that was of high enough quality to analyze the students' work. On average, the students took 40 minutes to complete the task.

Students' talk during the storyboard activity were transcribed and coded along several dimensions for a total of 2361 unique utterances. An utterance was defined as one breath group, or roughly one sentence. The science content codes identified honeybee structures, behaviors, and functions (Hmelo-Silver & Pfeffer, 2004). The other coding categories (described below) were developed from student talk using a grounded approach. Inter-coder reliability reached 85% agreement for 24% of all coded utterances.

To analyze the actual storyboards that the students created, a list was made of structures and behaviors based upon the pre- and post-interview-scoring rubric. We then amended this list by adding any additional structures or behaviors that were consistently mentioned in the storyboard videos, such as the fact that bees have 6 legs. Credit was given either for clearly identifiable visual features, or for references in the writing that students had included beneath the images. Two researchers coded each storyboard independently, and then compared their codes. Differences were negotiated until resolved so that the storyboard scores represent 100% agreement. These scores did not include a number of implicit features that were present in many of the post-storyboards, such as a pattern of showing the bee flying to the flower and then back to the hive. While these were technically accurate, it was difficult to attribute them to a deep understanding of the content. Also, we did not deduct points for inaccuracies such as a bee being labeled as a King bee, which primarily happened during the pre-test. Therefore, any reported gains are actually quite conservative.

## Results

### Storyboard Results

The mean score for a storyboard increased from 1.20 points to 5.55 points from the pre- to the post-activity (N=34 due to student absences). This increase was statistically significant: $t(34) = 11.115$, $p < .01$, two tailed. Some of the most common features that students were given credit for were the inclusion of 3 distinct body parts, the inclusion of 6 legs, and the presence of the proboscis. Figure 1 depicts representative bee drawings from the pre and post-test storyboards. The bee on the left would not have received any points while the bee on the right would have received points for having three body parts, a stinger, six legs, a clearly identifiable proboscis, a clear compound eye, and the pollen basket. Combined with the interview results referenced earlier, these findings suggest that not only did the students learn about how honeybees collect nectar over the course of this study, they also demonstrated this learning in their storyboards. With this learning as a backdrop, we now present the analysis of student talk while creating the storyboards. In particular, we will focus on several of the shifts in talk from the pre- to post-test condition as they coincided with these dramatic increases in what features the students incorporated into their storyboards.

Figure 1. An example of a bee from the pre-test (left) and from the post-test (right).

### Representational Practice Results

The first analysis we performed was to determine how much of students' talk throughout this sequence of activities related to the content being studied (either science talk, or the representations themselves). Each utterance was then coded as belonging to one of four categories: Science and Representation, Instrumental and Procedural, Spelling and Heading, and Off-topic (see Table 1 for descriptions and frequencies).

Table 1: Student utterances while creating their storyboards

| | Science and Representation | Instrumental and Procedural | Spelling and Heading | Off-topic |
|---|---|---|---|---|
| Description | Talk about the science content (honeybees) or the representation. | Talk about the procedure and materials such as crayons, or the storyboard template | Sounding out letters, or asking someone how to spell words | Talk about lunch, clothes, parents, pop-culture, etc |
| Utterances (N=2361) | 869 (36.8%) | 802 (33.9%) | 211 (8.9%) | 479 (20.2%) |

While our analytic focus for the remainder of this paper will be on the 36.8% of the talk where the students discussed the actual science content of the curriculum unit, or their ongoing representations, it is important that we not dismiss the fact that there remains an additional 42.8% of talk which was on-topic, and central to the students' work. This is important because we do not want to give the impression that the students were silent except when addressing the content. Nor should one consider a noisy classroom, in which many students are discussing issues other than the content being covered, problematic. Rather, much of this talk may be central to students' ability to engage with their ongoing work. Furthermore, we often noted that even when silent or discussing off-topic issues, students were frequently seen drawing. In other words, despite being quite revealing about students' understanding of content, on-topic talk is not the only predictor of how effectively the students were creating their representations.

## What aspects of the science content do students discuss as they create their storyboards?

In order to better understand the relationship between students' social practices while representing and the content being studied, we examined the science content of their talk. We first coded the students' science talk in terms of whether it referred to the structures, behaviors, or functions of how honeybees collect nectar. Our hypothesis was that students would talk primarily about structures and behaviors as this mirrors previous findings that describe students' performance when examining complex systems (Hmelo-Silver & Pfeffer, 2004). Furthermore, we hypothesized that there would be a significant increase in student talk about these features during the post-test as the students learned the content. Note that utterances were only coded if they represented a unique idea within an episode. Episodes consisted of groups of utterances around a similar topic. In other words, if a student mentioned the proboscis, and then discussed it for 5 utterances with their peer, this was only coded once in order to avoid inflating our results. Some utterances were also coded simultaneously for a structure and a behavior, or for more than one structure. Finally, note that the percentages refer to percentages of on-topic utterances in each condition, and so do not sum to 100 as there were also on-topic utterances related to representations and other issues to be discussed below. Frequency counts and examples are presented in table 2.

Table 2a: Students' science talk across time.

| | Structure | Behavior | Function |
|---|---|---|---|
| Examples | Bee, hive, proboscis (e.g. "This is a bee hive.") | Searching for flower, getting nectar, performing a dance (e.g. "They're going to drink nectar from the flower.") | Survive the winter, collect nectar more quickly (e.g. "They do a dance to tell others where the honey is.") |
| Pre (N = 367) | 106 (28.8%) | 86 (23.4%) | 11 (3.0)% |
| Post (N=502) | 113 (22.5%) | 34 (7.9%) | 2 (0.4%) |

As can be seen in table 2a, the most common type of science talk related to structures, as hypothesized (28.8% in the pre- and 22.5% in the post-condition). However, we were surprised at the fact that the percentage of on-topic science talk decreased from the pre- to post-storyboard. This decrease was statistically significant for unique counts of structure ($\chi^2(1, N = 869) = 4.56$, p = .03), unique counts of behavior ($\chi^2(1, N = 841) = 40.90$, p < .001), and unique counts of function ($\chi^2(1, N = 869) = 9.71$, p = .02). We will return to this point after first presenting some follow-up analyses.

To assess the quality of this science talk we applied three code modifiers, marking each science content utterance for whether it was scientifically inaccurate or not, anatomically detailed or not, and topically extraneous or not for the given science prompt: how do honeybees collect nectar. Note that, as before, these codes represent unique ideas within a given episode, and so these counts are somewhat conservative and again do not add up to 100% (See Table 2b).

Table 2b: Quality of students' science talk across time.

| | Inaccurate | Detailed | Extraneous |
|---|---|---|---|
| Examples | Flowers contain honey, there is a king bee (e.g. "This is the king and queen bees.") | Head, thorax, abdomen, proboscis, pollen basket, (e.g. "The pollen basket's right there.") | House, trash-can, jetpack, chairs (e.g. "This is how I make a person in a chair.") |
| Pre (N=367) | 11 (3%) | 0 (0%) | 19 (5.2%) |
| Post (N=502) | 1 (0.2%) | 24 (4.8%) | 8 (1.6%) |
| $\chi^2(1, N = 869)$= | 12.18, p < .001 | 18.04, p < .001 | 9.04, p = .003 |

Table 2b presents a somewhat different picture of students' science talk as it moves from the pre-test to the post-test, and may begin to explain why students' discussion of structures decreased despite our prediction that they would increase. While the percentage of talk featuring science structures decreased from the pre- to post-activity, the talk in the post-activity contained significantly fewer inaccuracies, more details, and fewer extraneous features than talk in the pre-activity (See Table 2b). This may also partially explain why the storyboard products in the pre-activity often contained inaccuracies, minimal details, and extraneous features, whereas the storyboard products in the post-activity presented fewer inaccuracies and many more details. In short, students' talk in the post-test was far more refined, focusing on accurate and relevant details.

To further explore the role of science talk in the students' representational practices, we then addressed the question, *How do the patterns of student talk change with respect to audience?* Despite a concerted effort to limit teacher interaction with students during the storyboard activity, students often addressed the teachers or responded to very high-level questions from teachers or other adults with very specific information. The results are summarized in table 3.

Table 3: Students' science talk across audiences.

| Audience | Structure | Behavior | Function |
|---|---|---|---|
| Peer (N = 503) | 99 (19.7%) | 33 (6.5%) | 5 (1.0%) |
| Adult (N = 293) | 87 (29.6%) | 83 (28.3%) | 8 (2.7%) |
| Self (N = 68) | 30 (44.1%) | 9 (13.2%) | 0 (0%) |

Students were more likely to explicitly mention science structures in comments made out loud to their selves (44.1%) or to adults (29.6%), than when speaking to their peers (19.7%). Similarly, students were more likely to explicitly identify honeybee behaviors when talking out loud to their self (13.2%), or an adult (28.3%), than when conversing with a peer (6.5%). Further, students were more likely to discuss functions relevant to nectar collection when speaking with adults (2.7%), than with a peer (1.0%) or out loud to themselves (0%). This highlights the importance of the teacher in soliciting content-rich discussion from the students given that these predominated even when the teachers were intentionally limiting their discussion with the students.

## What aspects of the representations do the students discuss?

To answer this question, students' representational talk was coded in two specific ways. First, it was coded in terms of whether the students were discussing what to represent, how to represent it, or why it should be represented in that way. These three categories were designed to capture the nuances of students' representational talk, and whether it engaged students in simply discussing the features of their specific representations, the method of implementing those features, or the underlying motivations and reasons behind them (Danish, 2009a).

Our grounded approach generated two further coding categories: Unspecified and Progress. Unspecified comments present an indiscriminate or vague subject (e.g. "Look at this" or "I'm doing something cool") and Progress comments refer to the students' current state of completion (e.g. "I'm on my third one," or "I'm almost done"). With respect to unspecified comments, note that while we could frequently determine the feature that the students were addressing by looking at the video, it was unspecified whether they were referring to the inclusion of the feature (the What code) or how it was represented (the How code). Prior analyses (Danish, 2009a) indicate that students focus on "what" was represented when drawing (as opposed to skits or simulations where students are more likely to discuss the "how" or "why" of their representational choices). Our hypothesis based on prior analyses is that students will primarily focus on the "what" features of representations during the pre-activity, but will shift to discussing the "how" and "why" of their storyboards more often in the post-activity. As predicted, the majority of students' talk in the pre-activity related to "what" was represented ($\chi^2(1, N = 869) = 47.13$, $p < .001$), whereas in the post-activity this talk shifted significantly to the "how" ($\chi^2(1, N = 869) = 25.47$, $p < .001$) and "why" ($\chi^2(1, N = 869) = 28.76$, $p < .001$) of the representations (see Table 4).

Table 4: Students' representational talk across time.

| | What | How | Why | Unspecified | Progress | Other |
|---|---|---|---|---|---|---|
| Pre (N = 367) | 232 (63.2%) | 27 (7.3%) | 13 (4.3)% | 35 (9.5%) | 6 (1.6%) | 46 (12.5%) |
| Post (N=502) | 199 (39.6%) | 98 (19.5%) | 73 (14.5%) | 45 (9.0%) | 36 (7.1%) | 49 (9.7%) |

As with students' science discourse, our goal in this analysis was to move beyond simply documenting the content of students' talk to capture the role that it played in their practices. Therefore, we further coded students' representational talk in terms of how it supported their interactions. Specifically, similar to prior work (Danish & Enyedy, 2007) we noticed that students typically engaged in assessing or critiquing each other's work. Therefore, we coded students' utterances in terms of the various ways that assessments played a role in shaping the students interactions: (a) Assessment-Seeking, (b) Assessment-Giving, (c) Assessment-Warranting, and (d) Assessment-Responding (see Table 5 for descriptions, examples, and frequency counts). Note that while we are reporting incidences of student talk in table 5, these comments often played a role in shaping students' representations, causing them to add, remove, or change the features of their drawings in response to the issues raised, even when the student making the change was not directly addressed (for similar examples see Danish & Enyedy, 2007).

Our statistical analysis reveals that students are significantly more likely to engage in assessment related interactions during the post-activity than in the pre-activity for our first three coding categories: (a) Assessment-Seeking ($\chi^2(1, N = 869) = 3.88$, $p = .048$), (b) Assessment-Giving ($\chi^2(1, N = 869) = 18.41$, $p < .001$), and (c) Assessment-Warranting ($\chi^2(1, N = 869) = 6.58$, $p = .010$). Assessment-Responding increased marginally from pre- to post-activity ($\chi^2(1, N = 869) = 3.01$, $p < .082$).

Table 5: Students' evaluative interactional moves over time.

| | Assessment-Seeking | Assessment-Giving | Assessment-Warranting | Assessment-Responding |
|---|---|---|---|---|
| Description | Directly requests evaluation of their storyboard | Explicitly appraises their storyboard or their peer's. | Augments their appraisal with a reason or example | Replies to the given feedback |
| Example | "What do you think of this?" | "it doesn't have wings," or "you forgot legs." | "the stinger is suppose to be a straight line" | "I forgot to include it" or "I'm waiting to draw it" |
| Pre (N=367) | 2 (0.5) | 6 (1.6%) | 2 (0.5%) | 4 (1.1%) |
| Post (N=502) | 11 (2.2%) | 42 (8.4%) | 15 (2.9%) | 14 (2.7%) |

## Discussion

This study aimed to document and describe the development of a group of kindergarten and first-grade students' representational practices as they related to drawing storyboards about honeybees collecting nectar. To highlight the change in students' practices, they were examined in a similar context (creating a storyboard) both before and after having learned about the content (how honeybees collect nectar). Our goal in doing so was to examine how the practices changed over time despite the fact that the task was, ostensibly, the same. We do not claim that the participants experienced the practice as the same at both time points. In fact, our theoretical assumption is that students' knowledge of the content and representing, and their shared practices for representing have shifted, which in turn shifts the way that they engage in representing at these two time points. In other words, the students learned new practices—which include an understanding of the content, how to represent it, and how to assess representations of it—and those new practices influenced how the students responded to the stated task of showing how bees get food using a storyboard.

Our results reveal that students' representational practices include a great deal of time talking about the task requirements (instrumental talk), some off-topic talk, as well as discussions of the content and how it should be represented. Initially, we were surprised to find that the students' talk included a higher percentage of unique ideas regarding science-related structures and behavior in the pre-activity than in the post-activity. However, further analyses revealed that those ideas were less accurate, less detailed, and were more likely to include non-essential structures (e.g. ladybug, house) to the science being represented. Furthermore, students were more likely to request and give assessments during the post-activity than during the pre-activity. These are important changes in representational practices that cannot be derived simply from counting the amount of structure and behavior talk.

A practice-oriented, and heavily contextualized analysis of students' representational acts defies, by definition, generalization of the specific findings. In other words, we would not expect to see this same pattern of talk in every kindergarten and first-grade classroom. However, we believe that some aspects of this analysis are, in fact, crucial to understanding and supporting other students and contexts. First, we have shown that there may be some value in examining the set of students' representational practices as they change over time instead of simply focusing on how they relate to the microgenesis of specific content knowledge as prior studies have. Had we not taken such steps, we might have overlooked the fact that students actually talk less about the content as they continue to master it—an important finding to consider when engaging in formative assessment of students' ideas through representational activities. This kind of analysis is supported, in part, by examining tasks that appear to be similar, rather than attempting to examine a shift in practices and context at the same time. Second, we hope that by documenting the breadth of student talk as they created their representations, we have helped practitioners and other researchers to recognize the fact that even students who have moved through a curriculum in a rather successful manner do not limit themselves to discussing the content accurately and consistently. However, one shift in student practices that appears to help correct for student mistakes and persistent misconceptions is the increase in critiques and other forms of assessment followed by students adjusting their representations accordingly. Future work should look more closely at the role of such critiques in supporting students as they create representations, and examine processes for encouraging more of them.

In sum, we have attempted to provide a robust and realistic picture of the kinds of talk that students engaged in as they created representations. Our hope is that this is a starting point for additional studies of students' representational practices which aim to look beyond the microgenesis of specific content ideas to a full description of the relationship between students' representing, the content, the context, and how these all shift over time.

## Endnotes

(1) Cobb et al., actually separate this discussion into several nested levels of analysis. However, for brevity, we compress them into the two components for our present discussion.

## Acknowledgments

We would like to acknowledge the teachers and students who so kindly participated in this study. We would also like to thank Noel Enyedy, William Sandoval, Frederick Ericson, Charles Goodwin, and Candy Goodwin who provided feedback on early versions of this analysis. Portions of this work were supported by the Spencer Foundation Dissertation Fellowship, and by a UC Regents Dissertation Fellowship.

## References

Callaghan, T. C. (1999). Early Understanding and Production of Graphic Symbols. *Child Development, 70*(6), 1314-1324.

Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in Classroom Mathematical Practices. *The Journal of the Learning Sciences, 10*, 113-163.

Danish, J. A. (2009a). *BeeSign: A Computationally-Mediated Intervention to Examine K-1 Students' Representational Activities in the Context of Teaching Complex Systems Concepts.* Unpublished Dissertation, University of California at Los Angeles, Los Angeles.

Danish, J. A. (2009b). BeeSign: a Design Experiment to Teach Kindergarten and First Grade Students About Honeybees From a Complex Systems Perspective. Paper presented at the annual meeting of the American Educational Research Association.

Danish, J. A., & Enyedy, N. (2007). Negotiated Representational Mediators: How Young Children Decide What to Include in Their Science Representations. *Science Education, 91*(1), 1-35.

DeLoache, J. S., & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition, 2*(52), 83-110.

DiSessa, A. A. (2004). Meta-Representation: Native Competence and Targets for Instruction. *Cognition and Instruction, 22*(3), 293-331.

DiSessa, A. A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing Graphing: Meta-Representationsal Expertise in Children. *Journal of Mathematical Behavior, 10*, 117-160.

Hall, R. (1996). Representation as Shared Activity: Situated Cognition and Dewey's Cartography of Experience. *Journal of the Learning Sciences, 5*(3), 209-238.

Hall, R., & Rubin, A. (1998). There's five little notches in here: Dilemmas in teaching and learning the conventional structure of rate. In J. G. Greeno & S. V. Goldman (Eds.), *Thinking practices in mathematics and science learning* (pp. 189-236). Mahwah, NJ: Lawrence Erlbaum Associates.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. . *Cognitive Science, 28*(1), 127-138.

Latour, B. (1988). Drawing Things Together. In M. Lynch & S. Woolgar (Eds.), *Representation in Scientific Practice* (pp. 19-68). Cambridge MA: MIT Press.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

Lehrer, R., & Schauble, L. (2000). Developing Model-Based Reasoning in Mathematics and Science. *Journal of Applied Developmental Psychology, 21*(1), 39-48.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. In R. W. Damon, K. Lerner, A. Renninger & I. E. Sigel (Eds.), *Handbook of child psychology, 6th edition, (vol. 4)*. Hoboken, NJ: Wiley.

Lynch, M. (1988). The Externalized Retina: Selection and Mathematization in the Visual Documentation of Objects in the Life Sciences. *Human Studies, 11*, 201-234.

Roth, W.-M., & McGinn, M. K. (1998). Inscriptions: Toward a Theory of Representing as Social Practice. *Review of Educational Research, 68*(1), 35-59.

Sherin, B. (2000). How students invent representations of motion a genetic account. *Journal of Mathematical Behavior, 19*, 399-441.

# When Students Speak, Who Listens? Constructing Audience in Classroom Argumentation

Leema K. Berland, University of Texas, 1 Univ. Station, Austin, TX 78712, leema.berland@mail.utexas.edu
Andrea Forte, Drexel University,3141 Chestnut St, Philadelphia, PA 19104, andrea.forte@drexel.edu

**Abstract:** Does a speaker craft an argument carefully if no one will hear it? Does an engineer design structures without reflecting on the individuals for whom her creations will have meaning and utility? Learning sciences research often explores such creative activities as meaningful opportunities for learning—inherent in these acts of creation is the concept of audience. In this paper, we draw on the philosophical and science studies literature to further unpack the role of audience in argumentation and examine three different ways that educators have addressed the problem of creating an audience for student work. We discuss how these approaches can challenge existing classroom structures and present evidence from empirical studies that demonstrate some of the outcomes.

## Audience: What and Who is It?

> *…the gathering of those whom the speaker wants to influence by his or her arguments.*
> (Perelman, 1982, p. 14), emphasis is his.

The word "audience" often conjures images of a crowded theater or lecture hall—the "gathering" referred to in Perelman's above quote. In reality, audience might be dispersed across continents and across centuries or may only exist in the mind of a speaker, writer, or artist. Walter Ong stated that, for writers, audience itself is a fiction, invoked by the arguer in order to organize and contextualize her communication (Ong, 1975). In addition, he suggests that the audience itself responds to a text by adopting an expected role—for example, the role of reviewer or the role of student—and participates in constructing meaning rather than simply receiving it. Whether real or imagined, listening or reading, live or in absentia, audience plays an important role in defining the context in which argumentation takes place.

Audience is complex. Understanding the nature and impact of audience plays a central role in the vast literature on rhetoric and argumentation. Whether written or oral, arguments are situated in a rhetorical space that is constructed by both the audience and the speaker or writer. Tindale described how audience plays an active role in determining the nature of an argument:

> The rhetorical audience is not a passive consumer of arguments, as some logicians seem to think; it plays an *active* role in the argumentation. The nature of the audience sets the terms of the premises, which are formulated in light of theses accepted by those to be addressed. The audience contributes assumptions to the reasoning… And the audience can interact with the argumentation in the mind of the arguer or in dialogue with the arguer… (Tindale, 1999, pp. 85-86)

Argumentation, then, is not simply a matter of constructing a logically coherent series of statements, it is a socially situated event designed by the arguer to satisfy the demands of a particular context. Thus, understanding audience in the classroom is critical for understanding how constructing arguments can be an effective learning activity and how to best design such experiences for students.

In this paper, we explore the role of audience when fostering scientific argumentation in classrooms. To support our analysis of audience in the classroom, we draw not only on the science studies and rhetoric literature, but on four years of empirical classroom studies that examine argumentation in both spoken (Berland, 2008, submitted; Kuhn, Kenyon, & Reiser, 2006) and written forms (Forte & Bruckman, 2006; Forte & Bruckman, 2007; Forte & Bruckman, under review). These different contexts for argumentation give rise to different understandings of audience in the classroom. We begin by exploring the role of audience in argumentative discourse, we then move to examine the implications for classrooms.

## Argumentation and Learning: Audience Influences the Goals of the Argument

The first order of business between an arguer and her audience is agreeing on a goal. Is the audience to be persuaded? Is a compromise sought? Are the parties mutually engaged in solving a problem? The goal of the interaction must be established. Even when the audience is present, this is typically done implicitly; in the case of written communication, the arguer may invoke an audience based on her knowledge of communicative norms for a particular genre of writing. Determining the goal is an important first step because different goals result in different argument styles. In fact, philosophers have identified multiple forms of arguments. For example, Walton (1998), drawing from work such as Aristotle (1955) and Locke (1961 [1690]), has identified six possible argumentative dialogues each of which is "a goal-directed conventional framework in which two [or more] speech partners reason together in an orderly way….Each type of dialogue has distinctive goals as well as methods that are used by the participants to achieve these goals together" (p. 3). Work such as this suggests that argumentative discussions vary broadly based on the participants' goals. Table 1 both summarizes Walton's dialogue types (as he describes them) and identifies the instructional goals emphasized by each dialogue type.

Table 1: Walton's dialogue types and implied instructional goals (Walton, 1998)

| **Dialogue type** | **End result** | **Implied Instructional Goals** |
|---|---|---|
| Critical | Individuals are persuaded of a single claim | Criticizing counter-arguments |
| Inquiry | Claims are proven to be true or false | Collaborating to draw conclusions from premises/data |
| Negotiation | A "good deal" is reached | Bargaining and compromising |
| Information-seeking | An individual gains information | Asking questions and collecting information |
| Deliberation | A practical problem is solved | Debating possible courses of outcomes by predicting their outcome |
| Eristic | Opponents are defeated | Quarreling through aggressive tactics (including insults, emotional attacks etc.) |

As seen in Table 1, each of these dialogue types has a different end result and suggests different kinds of pedagogic goals. For example, the critical dialogue is used to persuade people and involves the goals of understanding and refuting counter-arguments, whereas the inquiry dialogue is a way to construct a single claim that is irrefutably supported by evidence and entails collaborating to use data in order to draw conclusions.

The first author has examined argumentation in science classrooms (Berland, submitted a) in which students engaged in whole class arguments in a way that aligned with the critical dialogue of Walton's scheme (1998). Each of these is characterized by the various discourse moves receiving different emphases. For example, in the critical dialogue, students were likely to negatively evaluate one another while mustering evidence to disprove counter claims, while in the information-seeking dialogue the students questioned one another but rarely compared across their contrary claims in order to acknowledge or reconcile their disagreements. In addition, the first author witnessed students engaging in a form of argumentative discourse that did not align with Walton's scheme. In particular, the students engaged in "information sharing" in which they presented previously constructed arguments without attending to the arguments of others or receiving feedback on their own. Jimenez-Aleixandre, Rodriguez, & Duschl (2000) describe these interactions by saying the students are "doing school" rather than "doing science." They are characterized by a focus on procedures, the teacher's expectations and school culture (i.e., expecting the answer include one concept that is predetermined). In these dialogues, the often-implicit goal is to quickly get an answer to the question at hand. In other words, the goal is to construct an acceptable claim.

This information sharing dialogue generally does not align with expectations for argumentation—that is probably why it is not accounted for by Walton's scheme. However, we include it here because it reveals the importance of having an authentic audience: In these interactions the implied audience is invariably the teacher as students attempt to meet the requirements of the assignment. An audience of classmates or an external body might help move these students from the their focus on teacher expectations to the strength of their arguments.

Features of these dialogue types also surface when students invoke an audience for written work. The second author has examined students' writing practices in high school science classes when they are asked to publish a real science resource on a wiki. Students who wrote on the wiki also adopted the dialogic style of "information sharing." At first glance, it seems unlikely that one would find other forms of dialog in a written performance—if one is writing about science rather than discussing it, then it seems natural to present finished arguments. Yet, social media like wikis support discursive forms of written interaction as collaborators deliberate over the content they wish to present as well as the form of their argument. Although inquiry and negotiation dialogues such as these are common on other wikis, like Wikipedia, this kind of dialogue was rarely observed among students and was not encouraged by teachers (Forte & Bruckman, 2007).

These studies also revealed that, even when writing in an online public venue, students frequently adopted traditional standards for written school genres. These standards assume that the student's written performance is ultimately intended to be assessed by the teacher, not to persuade or inform others and not as a starting point for collaborative inquiry. Performative standards motivate student engagement in information-sharing rather than more critical or inquiry dialogues. Moreover, in this case, the students saw information-sharing as being in competition with other more argumentative dialogue types: students frequently reported invoking an inexpert audience to help them organize their arguments, and several students explained in interviews that the teacher as audience competed with their potential readership online (Forte and Bruckman, in press). Ultimately, they attended to both audiences as they constructed their texts.

> What do people want to know about [this disorder]? That's kind of like what I was thinking about. Like what would I be interested in if I was trying to look at one of these sites? And I wouldn't say I'd be really interested in the biochemistry but since it has to be included, it shall be included. – Sara

Combining our empirical studies with the philosophical and science studies literature emphasizes the importance of argumentative goals—and the role of the audience in determining these goals. In particular, we see in the work of both authors that the teacher-audience often worked to disrupt the students' focus on goals that align more directly with argumentation and sensemaking. This suggests that one challenge facing classroom communities as they work to engage students in argumentative discourse is the creation of contexts in which students are able to engage in argumentative dialogues and the goals implied by them.

## Audience Influences How the Goals are Achieved

Beyond influencing the dialogue type—or argumentative goal—the audience influences the content of the interaction by impacting the criteria by which the argument will be judged. As stated by Tindale, "….depending on the circumstances, their [arguers'] arguments will seek different results and will use methods appropriate to the purpose of the discourse as well as to the audience to be influenced" (Tindale, 1999). In other words, arguers must change the content of their argument to meet the expectations of their audience. If this alignment does not occur, the argument will fail because the audience will judge it on different criteria then the author used.

In terms of scientific argumentation that occurs in classrooms, we have seen the students' criteria for evaluating arguments range from superficial (i.e., evaluating the appearance of the graphs; assessing arguments based on assumed expertise of the authors) to more content-based (i.e., aligning the claim to an answer in a text book) to scientific (i.e., examining the alignment between the claims and the evidence presented in defense of those claims). If the criteria that audience members use influence the content of the argument, then we would expect these different criteria to result in different arguments. For example, if the audience is focusing on superficial criteria, the arguers have little reason to articulate the evidence that supports their claims. Similarly, if the audience is focused on whether the claim makes sense or aligns with a textbook answer, the arguers have little reason to do more than identify outside sources that validate their claims.

The relationship between the expectations of the audience and the content of the argument is made most obvious in discussions in which the arguer's approach to supporting a claim differs from the expectations of the audience. For example, in an earlier study (Kuhn, et al., 2006) the first author observed a group of students arguing about their explanations regarding why the majority of Galapagos Finches died in the mid 1970s. One pair in this group (correctly) thought that a drought killed the finches' food (plant seeds) while the other pair believed it was torrential rains. The torrential rain pair defended their claim using logic and personal experience with drowning plants while the pair that believed it was a drought relied on evidence of the rainfall decreasing. After being unable to meet their demands for empirical evidence, the group that believed that torrential rains killed the plants revised their claim to align with the available evidence: a drought killed the birds' food. In interactions such as this, you see the substance of the argument changing to account for the audience's expectations. In this case, the claim changed because the arguers could not meet their audience members' demand for empirical evidence that supported the original claim.

The second author found that, when writing, audience awareness was a constant feature of students' process for constructing a text that would be acceptable to both their teacher and to a broader readership. In general, students adapted their writing to fulfill the perceived expectations of their Internet audience (Forte & Bruckman, in press). Some commented on the sense of responsibility they felt toward their readership: "It's like 'oh my gosh, I have a huge responsibility now' even if nobody actually uses this. It's still there, somebody *could* use it so everything has to be exactly right." Comments like this demonstrate the student authors attending to a non-teacher audience. Moreover, these different audiences for the students' work demand different kinds of information and different forms of presentation. In fact, as students reconstructed their writing process in interviews throughout the school year, they described different strategies for constructing texts and citing their sources depending on whether they were attending to their Internet or in-school audience (See Table 3).

Many students explained that, in order to present scientific information to a broad audience, they had to simplify the complex vocabulary they found in scientific journals so that their readers would understand their assertions. In order to do so, students reported that they needed to engage deeply with content and they frequently looked up additional explanations and definitions to get things right. One student described his process for researching his topic as 40% "collecting information. 60% was actually trying to understand what on earth are we talking about." Another noted that:

> I had to take words and I mean, I wanted to make it readable for people too, because I knew other people were going to look at this, so I was not going to use, like, huge words, just kind of make it simple. - Lina

This same student also described consciously choosing to follow her teacher's suggestions even when they conflicted with her sense of the Internet audience's needs:

> Well, 'causes.' I was like, well, if someone was to look at this, it would be like 'causes of the disease'… and then Dr. Baker was like, 'Well, you are supposed to find the *etiology* of the disease.' I was like, 'Okay, I am going to use the word 'etiology'. - Lina

In this example, Lina ended up mirroring her teacher in order to perform for assessment purposes. The extra work that students had to do to develop their ideas for a less knowledgeable audience supports and explains Gunel, Hand, & McDermott's finding that writing science for the teacher is associated with lower post-test scores than writing for peers and younger students (2009). They similarly found that students tended to use "big words" to impress their teacher on the assumption that the teacher would already understand them.

Table 3: Student Strategies for Constructing a Science Text on a Wiki (Forte & Bruckman, in press)

<table>
<tr><th></th><th>Teacher as Audience<br>Goal: Meet Assessment Criteria</th><th>Broad Internet Audience<br>Goal: Provide Credible Science Resource</th></tr>
<tr><td>Crafting Content</td><td rowspan="2">• Follow directions<br>• Look at others' work<br>• Apply conventions from other classes<br>• Seek feedback</td><td>• Simplify scientific language<br>• Invoke an inexpert reader<br>• Use existing sources as a model<br>• Reflect on own experiences as a reader</td></tr>
<tr><td>Sourcing Content</td><td>• Same strategies, different rationale:<br>o Ensure credibility<br>o Defer responsibility</td></tr>
</table>

Beyond the clarity of their explanations, these students placed a different emphasis on citation as a criterion for successful arguments when thinking about their external audience. In fact, as these students wrote online for a broad audience, some explained that citation was an important part of presenting their work in public, because of the responsibility to their audience:

> All the information I put up had sources, had everything, had a credible background to it. So I think if it's going to be so open for other people to use, your work should be credible. – Jerry

The work of the first and second authors demonstrates that the students' sense of their audience influences the content and justification of their arguments. Moreover, the second author's work revealing that the students' shifted their criteria as their attention shifted from a teacher-based audience to an external readership demonstrates the impact of the teacher on their argumentation. In particular, we see that, as with their argumentative goal, the criteria that students apply to their arguments differ when they work to satisfy an authentic audience who might be persuaded by their argument rather than a teacher who seeks to assess whether they achieved the desired answer.

## Implications for Designing Learning Experiences

The intertwined relationship between audience and student engagement with the science content suggests that a primary challenge when engaging in argumentative discourse in science classrooms is to *create situations in which students have an authentic audience for their arguments*. Moreover, the corollary to the philosophical focus on how the audience can shape the form and content of an argument is that the audience has a role in the argumentative discourse. That is, although a fictional audience might be invoked, an audience that is present should be providing feedback and engaging actively in the construction of the argument. From the perspective of the audience, this suggests that the listeners and readers have a key role in the argument process. Thus, the audience members must learn to play that role: They need to ask questions, evaluate arguments and offer alternatives. Whether in the classroom or in mediated online environments, these practices may need to be actively introduced and cultivated.

In order to serve as an authentic audience in argumentative discourse, students must see one another's ideas as worth responding to and must value the feedback provided by their peers in addition to that provided by their teacher. In an examination of power in a middle school science classroom, Cornelius and Herrenkohl

(2004) found that the typical "evaluation" phase of the triadic-dialog (Lemke, 1990) or IRE (Mehan, 1979) exchange can limit student ownership of ideas. Similarly, Tabak and Baumgartner (2004) found that triadic dialogues limit students' need and motivation to substantively engage with their classmates' understandings. Analyses such as these indicate that the prominence of teacher feedback in class discussions can limit students' ability to engage with one another's ideas because it removes their authority and motivation to evaluate and question them. In order for students to be audience members for one another's arguments, classroom norms must change to enable such interaction.

In addition, if students are to act as an audience for one another, they must reach agreement about the goal of their interaction. As described in the first section, this agreement is often achieved implicitly and is restricted by the context and content of an argument; however, traditional school norms may conflict with goals of specific forms of discourse such as scientific argumentation. In an analysis of classroom communities' adaptations of scientific argumentation, the first author and colleagues (Berland, submitted a) found that each classroom adapted the discourse practice differently. Moreover, the students in that study never explicitly discussed their goals with one another—instead the goal was made clear in the students' interactions. Their apparent agreement in argumentative goal could be attributed to their existing classroom culture (Berland, submitted b). That is, the goals with which they argued were similar to the goals of their non-argumentative, more typical class discussions. For example, in the class that argued as a form of critical dialogue, the goal was to critique counter arguments in order to prove that they knew the right answer. During non-argumentative class discussions, these students also seemed focused on demonstrating that they knew the right answer. Similarly, the class that focused on information seeking during the class argument was frequently engaged in similar interactions with their teacher when not arguing. This work demonstrates that the goals that typically govern classroom practices influence how students engage in argumentation. Thus, creating an environment in which it is sensible for students to engage with their audience to negotiate goals and criteria for meeting those goals involves a transformation of social norms in the classroom.

## Transforming the Classroom Norms

Learning sciences research has a strong tradition of exploring transformative social arrangements in the classroom that cast students as responsible, generative participants in their educational experiences. For example, the literature on knowledge building communities suggests that students can function much like a community of scientists by proposing, debating and building on one another's ideas to further the knowledge of the class as a whole (Scardamalia & Bereiter, 1991). To support this goal, the software Knowledge Forum helps students identify possible discourse moves in a knowledge building exchange. Developing such a community requires a fundamental cultural shift in schools from what Scardamalia and Bereiter call a first-order learning environment, in which there is a static set of knowledge to be learned, to a second-order learning environment in which the state of knowledge is always changing and participants must continuously adapt to the ideas and suggestions of their peers.

Brown and Campione (1996) addressed this challenge by explicitly transforming the goals and criteria of the classroom activities. They did this by making students responsible for learning the ideas under study and then creating situations in which students were accountable to one another for those understandings. These expectations helped create a "community of learners," in which the students had authentic reason and opportunity to engage with one another's ideas through scientific discourse practices. However, these solutions required changing the entire community. This transformation is difficult, and possibly unrealistic in traditional classroom settings.

The first author built on this work with a focus on scientific argumentation: she designed activities in which the explicit goal was for students to convince one another of their ideas such that they had to attend and respond to one another's ideas rather than the teacher's. However, engagement in the goal of persuasion seemed to limit student motivation or willingness to engage in sensemaking (Berland, submitted a). This is an example of how changing a portion of the classroom norms—in this case, the interaction patterns—without changing the other expectations—such as the goals—resulted in an incomplete adoption of the scientific practice. This reinforces the systemic requirements put forth by the Brown and Campione (1996) and Scardamalia and Bereiter (1996) studies, but returns us to the question of how we can foster these sorts of interactions in more traditional classrooms.

Looking across the literature, we have identified three approaches to addressing this challenge.

1. Introducing an outside audience
2. Recasting student roles
3. Creating a fake audience

Each of these approaches has implicitly or explicitly created situations in which students had an authentic audience with whom they could negotiate the goals and criteria for their discussions. In the following sections, we will draw on the literature and our own empirical work to exemplify each of the three approaches.

## Approach 1: Introducing an Outside Audience

Science fairs, band concerts, and student journals are all examples of traditional ways that educators have sought to "break down" the barriers between the classroom and outside communities in order to introduce an external audience for student work. Recently social media and "Web 2.0" have brought about new opportunities for student interactions with outsiders (Ellison & Wu, 2008: Wheeler, Yeomans, & Wheeler, 2008). In addition, learning sciences researchers have a relatively long tradition of exploring the potential for students to interact with distant peers, mentors, and subject matter experts via the Internet. Kids as Global Scientists (KGS)/Biokids (Songer, 1996), CoVis Mentor Database (O'Neill & Gomez, 1998), and Bos and Krajcik's online writing research (1998) are all examples of early learning sciences projects that created opportunities for students to engage with and present ideas to distant audiences through networked technology.

Because these interactions are mediated through educational software, the goals of discourse and the criteria for achieving these goals may be influenced by the design of the communication environments themselves. For example, in an implementation of KGS, students used an interface that structured activities by allowing them to see weather data, ask questions of remote peers, and answer questions from remote peers and local graduate student mentors (Songer, 1996). This configuration supported an "inquiry" style dialogue (Walton, 1998) by encouraging specific kinds of interactions such as reviewing data and asking questions. In other cases, such as in the work of the second author, who studied students writing a science resource on a wiki, the communication tools themselves are relatively versatile and lend little structure to the interaction. In such cases, it is critical to structure learning activities in such a way that students have a model of how they are expected to engage in with their readership in order to develop the goal and criteria for those interactions (Forte & Bruckman, in press). This may be done by explicitly structuring assignments, or, in exceptional cases such as the Math Forum project, it may be done through the development of an online community with cultural norms and expectations for social interactions among members (Renninger, Shumar, Barab, Kling, & Gray, 2004). This ideal can be difficult to achieve when both students and teachers can easily fall back on "default" classroom interaction styles that place the teacher back in the position of primary audience.

As discussed above, if expected to join in a particular mode of discourse, audience members also need to understand their role in the interaction. This can be a particularly challenging aspect of introducing an outside audience—engaging in specific forms of argumentation and providing useful feedback is not always easy. Even when the interactive environment helps structure classroom interactions, it can be difficult for outsiders to know what is expected and how to take on the role of audience productively. Songer (1996) provides an example of effective mentoring in which a KGS mentor "seeds ideas" among students; they note that mentors were not always so effective and were coached to engage with students in specific ways. O'Neill and Gomez (1998) developed the CoVis Mentor Database based on their research with science teachers who orchestrated online mentoring experiences for their students and likewise note that both mentors and students "needed substantial guidance concerning the kind of help they should provide and expect" (p. 327).

## Approach 2: Recasting Student Roles

In the previous section, we discussed the use of an external audience to help transform classroom interactions. Another approach to changing the experience of audience is to recast students as audiences for one another. This can be done in more structured ways, such as assigning explicit roles, or less structured ways that encourage student engagement with one another's ideas in a more or less spontaneous fashion. Herrenkohl and Guerra (1998) took a highly structured approach in a study of participant structures in argumentation among elementary school science students. They identified three distinct intellectual roles in argumentative discourse: predicting and theorizing, summarizing results, and relating evidence and results to theory and prediction. They assigned students to these roles and, in a comparison group, also assigned audience roles to students. They found that when audience roles were specified, the class engaged in more whole-class inquiry activities and students took more leading roles.

Other researchers have had success assigning students less explicit roles. For example, in a study with undergraduate students, Forte and Bruckman (2006) found that assigning students to be reviewers of one another's work created a situation in which students tailored arguments to convince their peers. After reviewing their peers' arguments, students revised position papers to be more convincing to an audience that might not share their point of view. Still another form of audience role assignment that can implicitly encourage specific discourse practices is asking students to role-play with one another. For example, Pitts and Edelson (2004) ask students to adopt the role of scientists and interact with one another as such to solve environmental science problems. Although discourse roles themselves are not made explicit, students draw on their understandings of what scientists are like and what they do to organize their interactions with one another. Finally, a still less-structured approach is to give students a common goal to encourage knowledge-building interactions, such as Peters and Slotta's (2009) studies of students who construct a science resource for one another online.

When discourse roles are assigned, interactions can be far more structured than in the first approach and the problem of agreeing on argumentative goals and criteria is less difficult to resolve. The assigned roles should support students in meeting argumentative goals, which are often made explicit upon assignment. In the first example above, Herrenkohl and Guerra (1998) made highly specified role assignments based on scientific discourse practices. In this case, the activity was sufficiently structured to circumvent confusion about the goal of discourse and the means of achieving this goal. In the case of Pitts and Edelson's (2004) role-playing scientists, or Peters and Slotta's (2009) wiki writers, discourse is less highly structured and students must draw on their understandings of scientific practices to come to an agreement on the best ways to pursue their goals.

### Approach 3: Creating a Fake Audience

When it comes to scientific reasoning and argumentation, encouraging students to role play and engage with a fake audience (and even to believe such an audience exists) has been used to motivate and organize learning activities and to connect science learning with students' experiences outside of school. Elaborate cover stories are used in problem-based and project based learning projects such as *Blueprint for Success*, which places students in the position of designing a playground for an imaginary architecture firm (Barron, et al., 1998). In the *Mission to Mars* project, students designed model rockets to learn about physical science (Petrosino, 1998). After observing that students did not engage in inquiry learning while building their rockets, Petrosino introduced the idea that an audience would evaluate their designs and found that the introduction of a critical audience helped students organize their own activities in a goal-directed fashion cited in (Petrosino, 1998 cited in Barron et. al., 1998).

Once again, advances in educational technology have brought about opportunities for innovative ways to connect students with audiences—in this case, imaginary audiences. Goal-Based Scenarios (GBS) are a form of problem-based learning that engages students in an imaginary problem scenario (Schank, Fano, Bell, & Jona, 1994). The educational software *Sickle Cell Counselor* and *Broadcast News* are GBS's that encourage learners to develop arguments based on a dataset and present them to fictitious audiences who provide feedback and questions to direct further inquiry. These audiences are not real audiences that can be affected and persuaded by learners, they are computational agents that provide "canned" responses designed to elicit reflection. Because the audience does not exist, agreeing on an argumentative goal and on the criteria for successfully reaching that goal is a matter of coaxing students into the "right" frame of mind.

## Discussion

Comparing various educational reform efforts reveals that learning scientists and science educators have long been working to create situations in which students work to engage audiences in an authentic fashion. Examining each of these designs in light of the philosophical and social sciences work on argumentation and audience suggests that a key challenge connecting these studies lies in the students' interpretation of the audience: are they constructing arguments or designing artifacts for their teacher to assess for accuracy or for an audience that they can hope to influence? Each of the studies discussed above has created ways to make the experience more authentic by creating a sense of audience that was external from the teacher.

This analysis has also revealed the importance of attending to whether and how the students negotiate their goals for their interactions and their criteria for success with their audience. Is this something that comes from the teacher or as a natural outcome of their interactions? The second author's work suggests that having a teacher-imposed goal and criteria may be in competition with the natural goals and criteria that emerge through the interactions. However, the teacher needs a way to influence these interactions in order to help students stay focused on activities and discussions that will help them move towards the learning goals. Our work suggests that creating this balance should be a focus of future work in supporting students as they engage in argumentation as a learning activity.

## References

Aristotle (1955). *On sophistical refutations* (E. S. Forster, Trans. Loeb classical library edition ed.). Cambridge, MA: Harvard University Press.

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., et al. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *Journal of the Learning Sciences, 7*(3), 271-310.

Berland, L. K. (2008). *Understanding the composite practice that forms when classrooms take up the practice of scientific argumentation.* Northwestern University, Evanston, IL.

Berland, L. K. (submitted a). How classroom communities make sense of the practice of scientific argumentation.

Berland, L.K. (submitted b). Explaining variation in how classroom communities adapt the practice of scientific argumentation.

Bos, N., & Krajcik, J. (1998). *Students' awareness of audience in web-published science writing*. Paper presented at the Annual Meeting of the American Educational Research Association.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. pp. 289-325). Mahwah, NJ: Lawrence Erlbaum Associates.

Cornelius, L. L., & Herrenkohl, L. R. (2004). Power in the classroom: How the classroom environment shapes students' relationships with each other and with concepts. *Cognition and Instruction, 22*(4), 467-498.

Ellison, N., & Wu, Y. (2008). Blogging in the classroom: a preliminary exploration of student attitudes and impact on comprehension. *Journal of Educational Multimedia and Hypermedia., 17*(1), 99-122.

Forte, A., & Bruckman, A. (2006). *From Wikipedia to the Classroom: Exploring online publication and learning.* Proceedings of the Int'l Conference of the Learning Sciences, Bloomington, IN, 182-188.

Forte, A., & Bruckman, A. (2007). *Constructing text: wiki as a toolkit for (collaborative?) learning.* Proceedings of the International Symposium on Wikis (WikiSym), Montreal, Canada, 31-42.

Forte, A., & Bruckman, A. (In Press). Citing, writing and participatory media: wikis as learning environments in the high school classroom. *International Journal of Learning and Media*.

Gunel, M., Hand, B., & McDermott, M. A. (2009). Writing for different audiences: Effects on high-school students' conceptual understanding of biology. *Learning and Instruction, 19*(4), 354-367.

Jimenez-Aleixandre, M. P., Rodriguez, B. A., & Duschl, R. A. (2000). "Doing the lesson" or "Doing science": Argument in high school genetics. *Science Education, 84*, 757-792.

Kuhn, L., Kenyon, L., & Reiser, B. J. (2006). *Fostering scientific argumentation by creating a need for students to attend to each other's claims and evidence.* Proceedings of the International Conference of the Learning Sciences, Bloomington, IN, 370-375.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing.

Locke, J. (1961 [1690]). An essay concerning human understanding. In J. Yolton, W. (Ed.). London: Dent.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.

O'Neill, D. K., & Gomez, L. M. (1998). *Sustaining mentoring relationships on-line.* Proceedings of the CSCW 98: ACM conference on computer-supported cooperative work, 325 - 334.

Ong, W. (1975). The writer's audience is always a fiction. *PMLA, 90*(1), 9-21.

Perelman, C. (1982). *The Realm of Rhetoric*. London: University of Notre Dame Press.

Peters, V. L., & Slotta, J. D. (2009). Scaffolding knowledge communities in the classroom: New opportunities in the Web 2.0 era. In M. J. Jacobson & P. Reimann (Eds.), *Designs for Learning Environments of the Future*: Springer-Verlag.

Petrosino, A. (1998). *The use of reflection and revision in hands-on experimental activities by at-risk children.* Vanderbilt University, Nashville, TN.

Pitts, V. M., & Edelson, D. C. (2004). *Role, goal, and activity: A framework for characterizing participation and engagement in project-based learning environments.* . Proceedings of the ICLS 2004: Sixth International Conference of the Learning Sciences,

Renninger, K. A., Shumar, W., Barab, S., Kling, R., & Gray, J. (2004). The centrality of culture and community to participant learning at and with The Math Forum. In S. Barab, R. Kling & J. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning* (pp. 181-209): Cambridge University Press.

Scardamalia, M., & Bereiter, C. (1991). Higher Levels of Agency for Children in Knowledge Building: A challenge for the design of new knowledge media. *Journal of the Learning Sciences, 1*(1), 37-68.

Scardamalia, M., & Bereiter, C. (1996). Computer Support for Knowledge-Building Communities. In T. Koschmann (Ed.), *CSCL: Theory and Practice of an Emerging Paradigm* (pp. 249-268). Mahwah, NJ: Lawrence Erlbaum Associates.

Schank, R., Fano, A., Bell, B., & Jona, M. (1994). The design of goal-based scenarios. *Journal of the Learning Sciences, 3*(4), 305-345.

Songer, N. (1996). Exploring learning opportunities in network-enhanced classrooms: A case of kids as global scientists. *Journal of the Learning Scieces, 5*(4), 297-327.

Tabak, I., & Baumgartner, E. (2004). The teacher as partner: Exploring participant structures, symmetry, and identity work in scaffolding. *Cognition and Instruction, 22*(4), 393-429.

Tindale, C. W. (1999). *Acts of arguing: a rhetorical model of argument*. New York: State University of New York Press.

Walton, D. (1998). *The New Dialectic: Conversational Contexts of Argument*. Toronto: University of Toronto Press.

Wheeler, S., Yeomans, P., & Wheeler, D. (2008). The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. *British Journal of Educational Technology, 39*(6), 987-995.

# "Ideas First" in Collaborative Second Language (L2) Writing: An Exploratory Study

Yun Wen, Wenli Chen, Chee-Kit Looi
Learning Sciences Lab, National Institute of Education, 1, Nanyang Walk, Singapore 637616
Email: yun.wen@nie.edu.sg, wenli.chen@nie.edu.sg, cheekit.looi@nie.edu.sg

**Abstract:** Drawing from sociocultural perspectives, this paper argues for the necessity of distinguishing the activity of "brainstorming ideas" from the activity of "generating words/phrases" in collaborative L2 process writing. We designed a series of Chinese writing lessons based on the online collaborative learning software GroupScribbles (GS) to explore students' participation and performance with "ideas first" versus "words/phrase first" activity designs. Quantitative data collected from an analytical tool reveal that students engaged more actively in the "ideas first" collaborative L2 writing than in the "words/phrases first" approach. Our research findings suggest that "ideas first" supported by the affordances of the GS tools a viable approach for collaborative L2 writing.

## Introduction

Collaborative learning has been considered to be one of the effective instructional strategies in language learning. It "has a 'social constructivist' philosophical base, which views learning as construction of knowledge within a social context and which therefore encourages acculturation of individuals into a learning community" (Oxford, 1997, p 443). As an approach based on this notion, collaborative second language (L2) writing is a recursive, bottom-up process that requires participants to collaboratively contribute words/phrases or ideas, and eventually compose their compositions in individuals or groups. Although brainstorming as the fundamental pre-writing activity has been emphasized in various models of the writing process, almost all the models are established based on work in first language (L1) writing. Brainstorming is considered as an idea-generating technique which is an enabler for creative thinking. When models incorporating brainstorming are utilized to guide L2 writing, there is a salient contradiction between L2 writers' ideas expressions and their limited target language proficiency. Some researchers propose the model for collaborative writing process with vocabulary first at the initial step (e.g., Wong, et. al., 2009). In practical collaborative L2 writing activities, the teacher typically encourages the students to write down as much as they are able to without foregrounding ideas or words/phases contributions. In fact, for L2 writers, their low proficiency of the target language often requires them to focus primarily on vocabulary and grammar, and hence hampers ideas generation and expression (Scott, 1996; Stapa & Majid, 2009).

Ideas could take the form of arguments or evidences for an argumentation; they could be inspirations or original thoughts for a narration as well. We use the term "Ideas first" to refer to a particular design of the writing activity which foregrounds the content of ideas rather than their expressions or representations in some language. Students are encouraged to express own ideas related to a general topic in multimodal ways. For Chinese language learners, not only Chinese characters, but Hanyu Pinyin (Chinese phonetics) and even drawings can be accepted. Figure 1 (left) shows examples of artifacts manifesting "ideas" in this type of activity for the topic of: "Adolescent smoking is an increasing serious problem; discuss reasons for that". Figure 1 (right) shows the artifacts manifesting in "words/phrases" for the same topic. The "words/phrases first" activity foregrounds and requires knowledge of the target language, thus emphasizing more on rhetorical structures and grammar. Students are encouraged to contribute Chinese idioms or proverbs related to the topic, or proper nouns or terms or vocabulary which can be used in the given context, or various adjectives or verb phrases helpful for composing more precise and vivid expressions. For a student with higher target language proficiency, she or he may more automatically think of words related to the topic or the context. On the contrary, low proficiency students have a high propensity for weak or no participation. Therefore, it seems that the "ideas first" pedagogy should be distinguished from the "words/phrases first" pedagogy, to enable active participation at the beginning stage of the L2 collaborative writing process.

We propose that "ideas first" is a more effective approach to engage L2 learners in collaborative writing activities. Our study sets out to investigate students' participation in the "ideas first" activities and "words/phrases first" activities respectively. It involves four L2 writing lessons with two different pedagogical designs: two writing lessons involve "ideas first" activities and another two lessons involve "vocabulary first" activities. All the collaborative activities are enabled on a software platform named GroupScribbles (GS), which can facilitate peer online communication and ideas/vocabularies brainstorming and sharing.

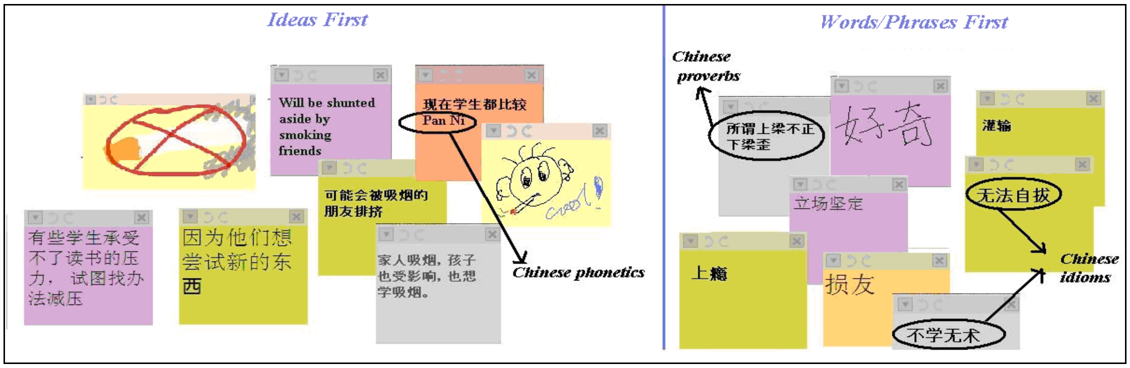



Figure 1. Examples of artifacts from “ideas first” vs. “words/phrases first” activities

## The Viability and Importance of “Ideas First”

Many psycholinguists or neuropsychologists hold the view that a person’s knowledge of vocabulary is stored in his or her mental lexicon (Harley, 2008), whereas ideas may be stored as concepts and images not yet formulated in words, and can be represented in any form of language. From this theoretical perspective, it seems that the capacity for ideas generation could be separate from language proficiency. When describing their classical model for cognitive process writing, Flower and Hayes (1981) argue that the process of converting ideas into words on paper is analogous to translating, demanding writers to put abstract thoughts through the technical aspects of writing such as grammatical structures. When cognitive load is not expended on vocabulary and language structure, the enthusiasm and capacity of L2 writers to engage in writing and group work participation could be enhanced. Hence, when “ideas first” approach is adopted, there can be more opportunities to exploit participants’ creative potential.

Traditionally, writing is regarded as a ‘‘non-linear, exploratory and generative process whereby writers discover and reformulate their ideas as they attempt to approximate meaning’’ (Zamel, 1983, p. 165), and ideas come from stored knowledge and experiences directly related to the writing topic. However, this traditional cognitive approach is too narrow in its understanding of social, historical and political contexts of writing (Prior, 2006). Today, with the sociocultural theory being more pervasive in the field of second language learning, there are more concerns over the role of social interaction in students’ cognitive development and language learning. A central hypothesis underpins the sociocultural theory of human development whereby higher-order functions develop out of the social interaction of an individual with the external social world (Tharp & Gallimore, 1988) that includes people, objects, and events in the environment. In this view, language as one of the most crucial human artifacts is thought to be socially constructed rather than internally intrinsic (Thorne, 2000). The participation metaphor of learning has been widely accepted, which refers to learning as a process of participation in shared learning activities and social process of knowledge construction (Sfard, 1998). Accordingly, rather than interpreted as internal mental processes solely by the individuals, language learning is viewed as a semiotic process attributable to participation in social activities (Block, 2003; Lantolf, 2000; Lantolf & Thorne, 2006). Therefore “ideas” that can be expressed and shared do play a bigger role and is critical to the creativeness of a learning community. The production and continual improvement of ideas that have value to a community is based on the premise that what the community accomplishes will be greater than the sum of individual contributions and part of broader cultural efforts (Scardamalia & Bereiter, 2003).

Collaborative writing should be seen beyond being just a mere instructional strategy with the obvious advantage that “two heads are better than one”. The collaborative writing process could be regarded as a process of meaning-making in which language learning and cognitive development occur in the sociocultural activities of participants’ community. This is aligned with the emphasis of Scardamalia and Bereiter’s (1987) process writing model that makes the transition from knowledge telling to knowledge transforming.

## An Intervention Study

### Participants

The study described here is part of a 3-year project to introduce rapid knowledge building practices to primary and secondary schools in Singapore (Looi, Chen & Ng 2009). We co-designed and implemented GS-based collaborative lesson activities with school teachers for students’ Science, Mathematics and Chinese language learning. In our work, we investigated useful pedagogies and teaching strategies that tap on the affordances of GS. This paper focuses on the collaborative L2 writing lessons with two types of pedagogical designs by using GS in secondary 3 (grade 9, 15-year old) normal Chinese language learning. In Singapore schools, English as the first language (L1) is the main teaching language in schools, whereas Chinese is taught as a second language (L2) for the Chinese ethnic students. In this study, the Chinese language class consists of 29 students. The students were organized into 7 groups (each group has 4 students, except a group with 5 students) in order to participate in the GS activities. Before the GS-based writing lessons, this class had been involved in our project for more than half a year (one GS lesson per week), and so they would have been enculturated with the GS environment and collaborative practices enabled by it (Looi, Chen, & Patton, submitted).

## GS interface and affordances

The lower pane is the user's personal work area, or "private board", with a virtual pad of fresh "scribble sheets" on which the user can draw or type (see Figure 2). Scribble sheets are the digital equivalent of small sticky notes that enable a lightweight scribbling to be made on the note. The essential feature of the GS is the combination of the private board where students can work individually and group boards or “public boards” where students can post the work and position relative to others’, view others’ work, and take items back to the private board for further elaboration. Students can choose any mode they feel comfortable to express their ideas through typing, writing by stylus or even drawing on the pad, and then post the pad onto the group or public board to share with others. A student can select any group board by clicking the board number on the right-top, and browse all other groups’ postings posted on the public board. GS hence promotes and facilitates intra- and inter-group sharing of ideas/information.

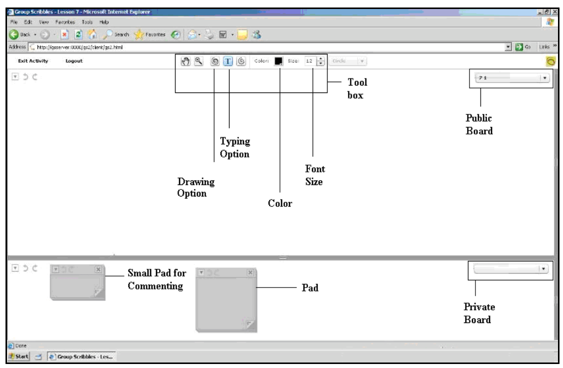



Figure 2. The user interface of GS with a two-paned window

GS is designed for students’ collaborative generation, collection and aggregation of ideas through a shared space based upon individual effort and social sharing of notes. Our research work seeks to tap the potential of GS for supporting L2 learning. In language learning, the task posed for collaborative activities does not focus on problem-solving, but it may be targeted towards enriching students’ vocabulary and proficiency in language expression, developing their thoughts and writing through cogitating with new words, vocabulary and sentence construction (Looi, Chen & Wen, 2009). When exploring the affordances of GS for Chinese language learning, we find that the affordance of “lightweight participation” fits L2 learners very well. For L2 learners, it is not only difficult for them to express themselves in complicate long sentences, but also difficult for them to extract main points from long paragraphs in a short period of time. By ‘lightweight’ we mean each student’s contribution or post can be composed (written, sketched, or typed) quickly in a small size of scribble sheet. Both fragments of ideas and diversified words/phrases can be represented in scribble sheets. They can be shared democratically, organized conveniently and improved continuously.

## Lesson design

This study examines the differences of “idea first” pedagogy and the “words/phrases first” pedagogy in students L2 collaborative writing. Four GS-based Chinese writing lessons (each of 70 minutes) were co-designed by the teacher and researchers. The topic for the lessons 1 and 2 was “why do Singaporeans feel discontented?”, and the task for lessons 3 and 4 was to make a story based on the context that “when you see your classmates Xiaowei and Xiaoli whispering at the school gate, what will you think of?” The topics are designed to be relevant to the students’ daily life.

In lesson 1, students were required to start collaborative L2 writing by first brainstorming ideas on the GS platform. In lesson 2, students first organized their ideas and then selected a few good ones to be further developed into paragraphs through GS. In the “ideas first” activity, they were reminded that any mode of expressing ideas is acceptable. Every student could have an equal opportunity to participate in collaborative writing, including not only generating and sharing ideas, but also improving these ideas by negotiating with others.

One week later in lesson 3, the students were asked to start collaborative L2 writing by generating words/phrases related to a new topic to explore story plots. In lesson 4, they wrote the main paragraphs with the collected words/phrases in lesson 3. In the "words/phrases" activity, students were encouraged to contribute suitable Chinese words/phrases that could be adopted directly in the final writing. In the process of interacting with others, students could enlarge their vocabularies and equip themselves with better understanding of the collected words/phrases.

In lessons 1 and 3, students were given examples about ideas on the story and words/phrases related to the topic respectively before the GS activities. For lesson 1, students might be able to come up with more ideas as they were not bogged down by the confusion between linguistic information and ideas on the topic (Scott, 1996). During each lesson, a short duration of time (around 20 minutes) was allocated for the students to do a "virtual gallery walk" via scanning the boards of other groups in GS. Each student could proffer comments or "borrow" or build on good ideas/words from other groups. After lessons 2 and 4, the teacher printed out each group's final products (GS screen capture) for the students who were then required to individually write the composition.

## Data Analysis

The analysis of this study includes two parts. In the first part, the GS analytic tool is deployed to probe the students' engagement levels within a lesson. This analytic tool is a program developed by us to extract quantitative information which serves as a proxy for the participation levels of each student in the GS activities. The second part focuses on examining group performance differences under the two pedagogical approaches.

Figure 3 and 4 show the students' participation in GS activities in lesson 1 ("ideas first" activity) and lesson 3 ("words first" activity) respectively. The horizontal axis represents the time of the class period using an interval of 6 minutes. The vertical axis is the aggregate count of all students' actions (including typing, drawing, moving postings etc) on the GS boards. When examining the students' GS actions in the first 6 minutes, we found that students had more participation in the GS based collaborative L2 writing in the "ideas first" activity than in the "words first" activity. There were a total of 10030 actions in lesson 1 and 9851 actions in lesson 3. The data perhaps does not mean that the class students posted a larger number of scribble sheets in the ideas first activity, but it implies that there were more students' participations or interactions via GS. They might contribute more actively, or click on other postings and modify the postings of their peers more frequently.

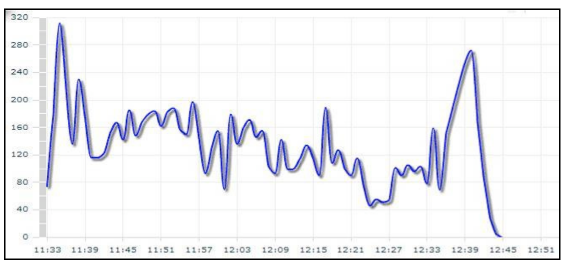



Figure 3. Count of whole class's actions on GS board in lesson 1 "ideas first" activity

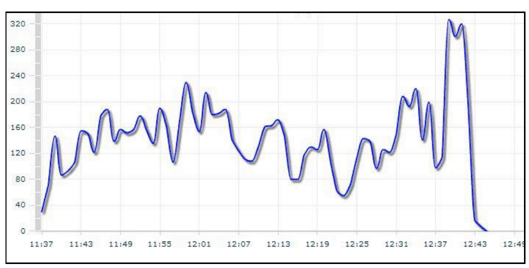



Figure 4. Count of whole class's actions on GS boards in lesson 3 "words first" activity

We further analyzed the data from two groups, namely, group 1 and group 6 (See table 1), to investigate different Chinese language proficiency groups' performances in the same activity. Both groups had a homogenous team composition with 4 students, and the team members were chosen based on their Chinese subject exam results before using GS (class average score was 57). The students' Chinese language proficiency in group 1 (with an average score of 51) is much lower than the students from group 6 (with an average score of 62).

Both groups reported good intra-group collaboration. In their post-lesson reflections, students from group 1 wrote that "my group can collaborate well and complete a task together…can communicate well…

correct each other mistakes and try to improve each other's Chinese". Likewise, students from group 6 considered their group as "a good one" and "a quite efficient group". Students' comments on their group collaboration are consistent with the judgment of the researchers based on the classroom observation.

Table 1. The information of two target groups

| Target Group | Group Configuration | | | Group culture of collaboration |
|---|---|---|---|---|
| | Division | Group members | Proficiency in Chinese | |
| Group 1 | homogenous | 4 (2 girls and 2 boys) | Low | Good |
| Group 6 | homogenous | 4 (3 girls and 1 boys) | High | Good |

Figure 5 compares the participation levels of two groups in the "ideas first" activity and the "words/phrases first" activity. We note that the difference in language proficiencies has impact on the participation levels of two groups. Group 1 student (an average of 20 actions per 6 minutes) seems to be more active in the "ideas first" activity than group 6 students (an average of 17 actions per 6 minutes) though group 1's members have lower language proficiencies than that of group 6. The difference in participation levels between the 2 groups in the "words/phrases first" activity was not prominent (both on the average 18 actions per 6 minutes). It seems that compared with students from high language proficiency group, low language proficiency group students were more active in "ideas first" activity. A plausible explanation is that they were less limited by language proficiency in "ideas first" activity since they were allowed to use drawings or their first language to express their ideas.

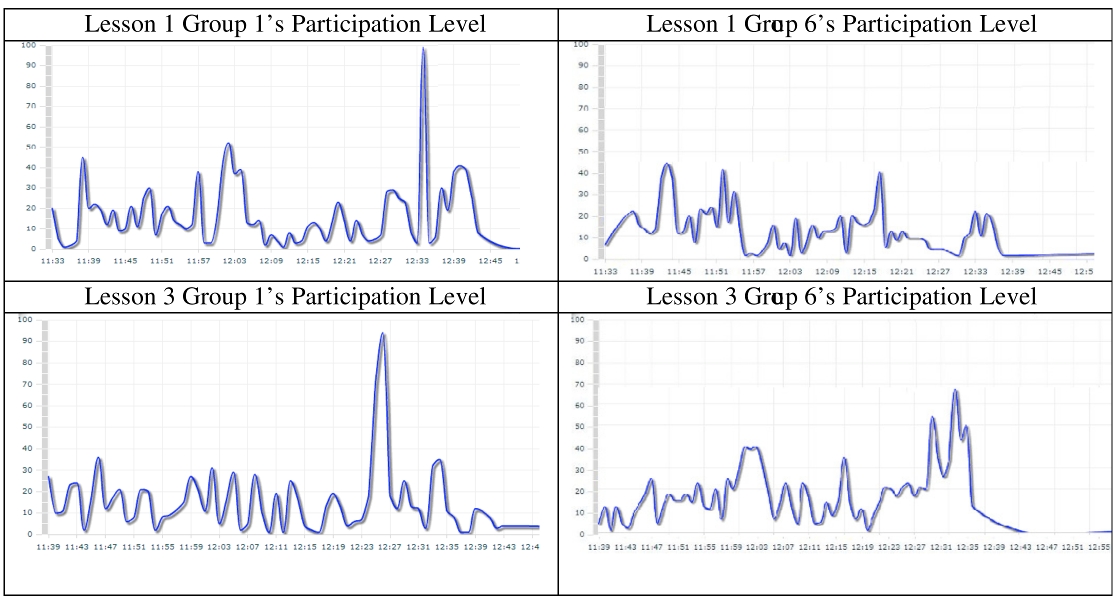



Figure 5. Group participation in "ideas first" activity *vs* "words/phrases first" activity

As shared by the teacher, the weaker ability groups typically had very low motivation to participate in Chinese writing activities. However, after the implementation of the "ideas first" activity, she said "I did not expect the weaker ability groups could complete the task well". The final GS artifacts of group 1 (Figure 6) were better than the artifacts from other higher ability groups in the "ideas first" activity in terms of richness of ideas and the logic of the ideas. The upper section of the screen shows three arguments for the topic "why do Singaporeans feel discontented?" (circled in dotted lines). The lower section shows three solutions to explain the issue at hand (highlighted by dotted rectangles). Other scattered postings were the comments given by other groups when they were visiting group 1's board. On the whole, after the activity of idea generation and modification in lesson 1, the final GS product of group 1 was well-organized, although some sentences were expressed in an awkward style.

The group selected specific individual ideas and developed them further through idea refinement and group negotiations into arguments why Singaporeans feel not contented. For example, the contribution of "税太高（too high tax rate）", referring to the sales tax, was first mooted by group 1 but was never been mentioned by other groups before the gallery work. In the activity of brainstorming ideas, a girl in the group first generated it in English, and then a boy translated it into Chinese. Later with the help of teacher, they posted "太多税要给

（too much taxes to pay）". Finally in lesson 2, they generated the argument which we highlighted by the dotted ellipse at the right-top of the Figure 6, namely, that "Singaporeans always worry about the high tax rate as high taxes made it less affordable to eke out a living." When challenged by other groups to explain why reducing the sales tax would make people feel contented, they clarified that "if the government reduces the tax, it becomes more affordable for the people to make purchases".

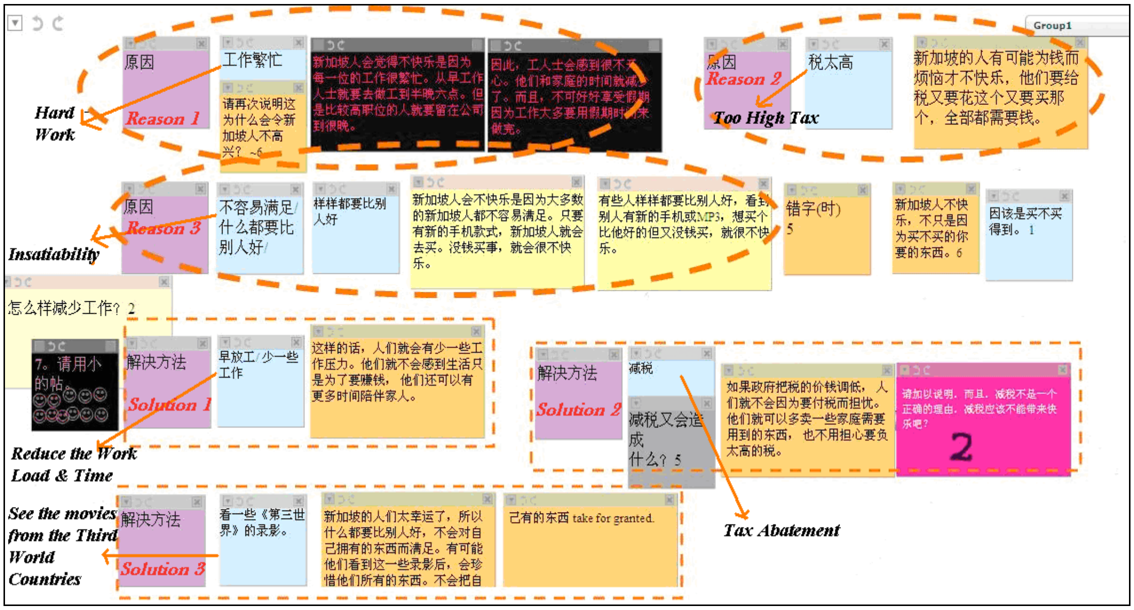



Figure 6. GS screenshot from Group 1 after the "ideas first" activity

In the "words/phrases first" activity, the weakness in language proficiency of group 1 became evident when they were generating words in lesson 1. Figure 7 displays the products of the two groups at the end of the words/phrases contribution phase. The left side of the figure shows the words and phrases generated by group 1, and the right side shows those from the higher ability group. For the benefit of the reader, we circled and translated some simple words posted by two groups, and highlighted comparatively complicated words and idioms within rectangles. Words and phrases from group 1 were much simpler than those from group 6. From a scan of the notes' colours (different students posted in different colours), we found that for both groups, the activity of posting words and phrases were dominated by one group member who had higher language proficiency in the group. In this type of activity, it was difficult to achieve democratic participation within the group collaboration as some students were restrained by their individual language proficiency. In lesson 4, during the process of extending words and phrases into paragraphs, little creativity could be identified from the students' work. This was despite the fact that teacher had constantly reminded the students: "I would like to see more creative stories… do not give me the same stories." Six groups of students wrote similar love stories which were rather cliché.

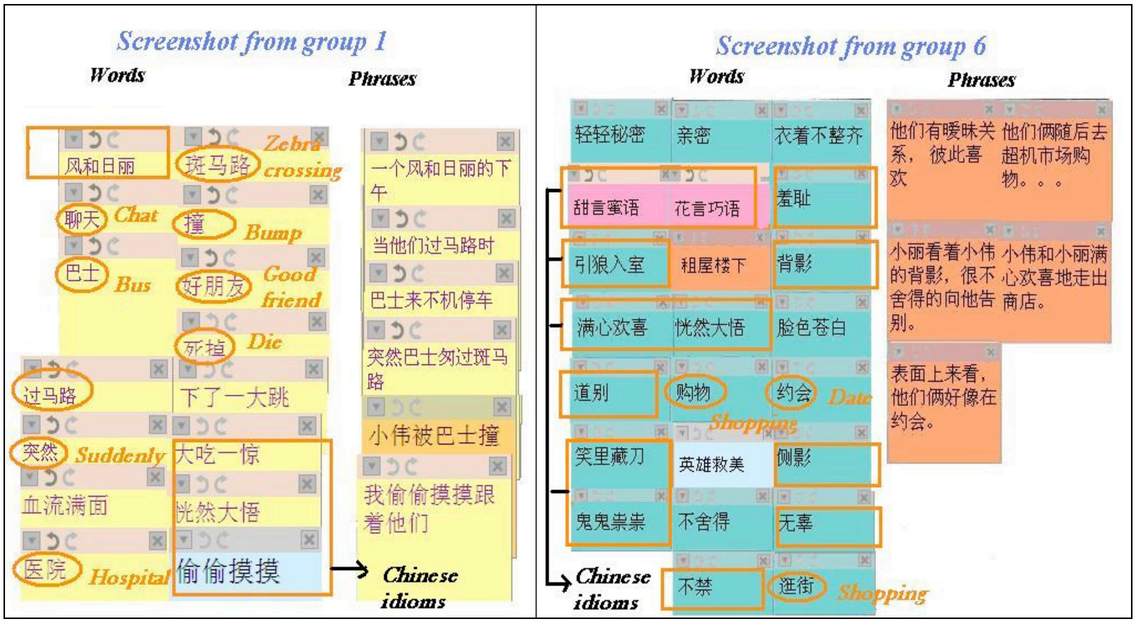



Figure 7. The GS screenshot of two groups in the "words/phrases first" activity

## Discussion & Conclusion

The collaborative L2 writing is a complex process which can be affected by various factors. This paper focuses research effort on the role of ideas generation in collaborative L2 writing. Two GS-based Chinese writing lessons, with the "ideas first" and "words/phrases first" activity designs, are compared. The findings suggest that the "ideas first" activity is a viable approach for collaborative L2 writing. Students especially those who do not have good language proficiency more actively participated in the collaborative L2 writing "ideas first" activity. GS artifacts arising from the two different learning activities also provide evidences that the "ideas first" activity encouraged more symmetric participation from students, and engendered more creativity in their writing

The affordances of GS, "multimodal expression" provides the students more modalities to engage in the "ideas first" activities, whereas in the "words/phrases first" the opportunity of participation is "reduced" to words, and which may generally result less contributions on the GS board. In addition, "lightweight participation", plays an important role in students' collaborative L2 writing. Due to the size limit of scribble sheet, students have to use brief and recapitulative phrases or sentences to express their ideas in the GS environment. It does not matter if the ideas are not mature. The smallness of scribble sheet encourages every individual to think and share actively, by emphasizing "ideas" rather than language proficiency. Furthermore, the students do not need to worry about the organizations of the ideas at the beginning. In a mature learning community, diversified forms of expression could be transformed and organized at the later stage of the learning activity. During this process, new ideas are generated, refined and improved continuously. As the GS activity supports inter-group interactions, ideas could be spread throughout the whole class. A number of studies (e.g., Lockhart & Ng, 1995; Nelson & Carson, 1998) have shown that when students are involved in L2 collaborative writing, they tend to edit the vocabulary or grammatical errors of their peers. Thus, initial representations of ideas do have the opportunities of being continuously polished, and that enables students to write better individual compositions eventually. On the contrary in the "words/phrases first" activity, the students have to rely much on their language proficiency. Even though L2 learners can participate in the group work anonymously, they may not be able to generate words/phrases effectively due to their low language proficiency. Thus students' symmetric participation in the learning activities may still be hindered when the writing activity starts from the words/phrases contribution.

This study is exploratory in nature. Our school-based research is still ongoing and we plan to explore further how the ideas or words spread and are improved upon within and across different student groups in the collaborative L2 writing activities.

## References


Block, D. (2003). *The social turn in second language acquisition* Washington, DC: Georgetown University Press.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*. *32* (4). 365-387.

Harley, T. A. (2008). *The psychology of language*. USA, New York: Psychology Press.
Lantolf, J. (2000). Introducing sociocultural theory. In J. Lantolf (Ed.), *Sociocultural Theory and Second Language Acquisition*, Chapter 1 (pp. 1-26). Oxford: Oxford University
Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. Oxford, UK: Oxford University Press.
Lockhart, C. and Ng, P. (1995). Analyzing talk in ESL peer response groups: Stances, functions and content, *Language learning 45* (4), pp. 605–655.
Looi, C.K., Chen, W., & Ng, F. K. (2009). Collaborative activities enabled by GroupScribbles (GS): An exploratory study of learning effectiveness. *Computers & Education* (2009), doi:10.1016/j.compedu.2009.07.003.
Looi, C.K., Chen, W., & Wen, Y. (2009). Exploring representational practices interactional moves in a CSCL Environment for Chinese Language Learning, *Proceedings of CSCL 2009*, ISLS.
Looi, C.K., Chen, W., & Patton, C. (submitted) Rapid collaborative knowledge building. Paper submitted to *Educational Technology*.
Nelson, G. L., & Carson, J. G. (1998). ESL students' perceptions of effectiveness in peer response groups, *Journal of Second Language Writing 7*(2), pp. 113–131.
Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. The Modern Language Journal, 81(iv), 443-458.
Prior, P. (2006). A sociocultural theory of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 54–66). New York: The Guilford Press.
Scott, V.M. (1996). *Rethinking Foreign Language Writing*. Boston: Heinle & Heinle.
Scardamalia, M., & Bereiter, C. (2003). Knowledge Building. In *In Encyclopedia of Education*, Second Edition (pp. 1370-1373). USA: New York: Macmillan Reference
Scardamalia, M. & Bereiter, C. (1987) Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.) *Advances in Applied Psycholinguistics*, Vol. 2. pp. 142-175
Sfard, A. (1998). On two metaphors for learning on the danger of choosing just one. *Educational Research, 27* (2), 4-13.
Stapa, S. H., & Majid, A. H. A. (2009). The use of first language in developing ideas in second language writing. *European Journal of Social Sciences, 7*(4), 41-48.
Tharp, R. G., & Gallimore, R. (1988). *Rousing Minds to Life: Teaching, Learning, and Schooling in Social Context. Cambridge*: Cambridge University Press.
Thorne, S. L. (2000). Second language acquisition theory and the truth(s) about relativity In J. Lantolf (Ed.), *Sociocultural Theory and Second Language Learning* (pp. 219-243): Oxford: Oxford University Press.
Wong, L. H., Chin, C. K., Chen, W., & Gao, P. (2009). V.S.P.O.W.: An innovative collaborative writing approach to improve Chinese as L2 pupils' linguistic skills. *Proceedings of CSCL 2009*, Rhodes Island, Greece.
Zuengler, J., & Miller, E. R. (2006). Cognitive and sociocultural perspectives: Two parallel SLA worlds? *TESOL Quarterly, 40* (1). 35-59.

## Acknowledgements

This material is based on the work supported by the National Science Foundation (United States) under Grants 0427783 and 0713711 and by the National Research Foundation (Singapore) under Grant NRF2007-IDM003-MOE-001. The GS software is available as a free download by SRI International and the National Institute of Education (Singapore). We are also grateful to Whitley Secondary School (Singapore) for collaborating with us on this research.

# Large Scale Analysis of Student Workbooks: What Can We Learn About Learning?

Nicole Shechtman, SRI International, nicole.shechtman@sri.com
Jeremy Roschelle, SRI International, jeremy.roschelle@sri.com

**Abstract:** As Learning Science–based innovations are studied at scale, traditional Learning Sciences methods such as video analysis and classroom observation become impractical. Yet Learning Scientists want to know more about student misconceptions, the connections between writing and conceptual understanding, and classroom practices than pretests and posttests can reveal. In this paper, we discuss an exploratory posthoc analysis of 765 workbooks from 48 classrooms that implemented SimCalc. These classrooms participated in a large-scale experiment in which we found that students learned more advanced mathematics in classrooms that implemented SimCalc. A team of three master teachers coded the workbooks for completeness, correctness, and other impressions. We found characteristics of students' work that predict gain scores, including the first evidence that classic SimCalc activities—writing stories and drawing graphs about motions—impact student learning. We discuss potential implications for large-scale Learning Sciences research.

## Introduction

Although much Learning Science research is conducted at a small scale (e.g., within a handful of classrooms per research study), most Learning Scientists want to have a large impact on potentially hundreds or thousands of schools. When going from one scale to another, more variance will inevitably occur. Classrooms are complex places and differ in myriad ways. Some variations in classrooms and in implementation may have dramatic consequences for the effectiveness of a design that was successful in smaller scale trials. Ann Brown, for example, famously highlighted the potential for "lethal mutations" that could undermine the effectiveness of an otherwise promising innovation (Brown & Campione, 1996). Thus, a key part of the scaling-up process is research on implementation (e.g., Cohen & Hill, 2001)—seeking to understand what kinds of variation in implementation occur and which variations matter for desired outcomes.

Research on implementation can entail the collection and analysis of a variety of different types of data, including video recordings, field observations, teacher logs, surveys, and artifacts from the classroom. At the scale considered in this work, involving implementation in 48 classrooms in the same school year, researchers face difficult tradeoff decisions in choosing what data to collect. One key tradeoff is between the *cost* of collecting data and the *quality of inferences* the resulting data can support. For example, video recordings are very expensive to collect and analyze but support in-depth analysis of a type that many Learning Scientists prefer. It can be much less expensive to collect and analyze a survey, but inferences from such data may be more circumscribed.

In this paper, we consider the potential value of collecting student workbooks in implementation research at scale. Student workbooks provide a potentially useful tradeoff. On the one hand, they are quite inexpensive to collect; we have found that teachers are willing to put them in a prepaid mailer, eliminating the need for a researcher to visit the classroom. On the other hand, workbooks directly exhibit student work, which can be a rich source of insights into students' misconceptions, for example. We report what we were able to learn from an exploratory, posthoc analysis of 765 workbooks.

We consider two kinds of research questions, some that are general to the method and some that are specific to the SimCalc intervention.

We asked three *method questions*:

1. How long does it take to code 48 classroom sets of workbooks for a 2–3 week intervention?
2. Can suitable interrater reliability be achieved?
3. Could raters accurately detect indicators of high-achieving and low-achieving classrooms?

With regard to the SimCalc intervention, the program developers espoused the beliefs that students need to be actively engaged in using the materials in order to learn (e.g., rather than just watching a teacher lecture with the materials). It is difficult to collect evidence of active engagement of all students in 48 classrooms. However, workbook completion does reveal the extent to which students were engaged in doing the work, especially because the curricular workbooks required students to use the software in order to answer various questions. Furthermore, because students drew graphs and wrote stories in their workbooks, we were able to consider the impact of student engagement with these forms of representation on learning.

Consequently, we asked an additional two *research questions*:

4. Does the completeness of student workbooks, as a proxy for students' active engagement in doing mathematics during the course of the replacement unit, predict student learning gains?

5. Does the correctness of details in student explanations, drawings of graphs, and telling of stories about graphs predict student learning gains of advanced mathematical concepts?

In addition to a specific focus on these research questions, we will report some additional anecdotal insights about what we learned from the analysis of workbooks.

## Background: Analyses of Student Work

In some ways, analysis of student work is ubiquitous in the Learning Sciences. Many published articles show artifacts from the classroom—including students drawings, calculations, graphs, and tables—to provide evidence for their arguments. Student work is often useful, for example, for identifying student misconceptions in graphing (Clement, 1989). However, most analyses are in the context of case studies. Few involve quantitative comparisons of students' prior knowledge and learning outcomes.

One notable use of student work in larger scale research was in the work of the Consortium on Chicago School Research, which collected samples of teacher assignments and student work in 12 Chicago schools at three grade levels and both in mathematics and in language arts (Newman, Lopez, & Bryk, 1998). The researchers found that when teachers gave more challenging assignments (i.e., that required a higher level of intellectual work), students were more likely to exhibit high level work and to have higher achievement outcomes.

Analysis of student work is also featured as a method of teacher professional development and improvement of instruction. For example, in Cognitively Guided Instruction, the entry point for teachers into students' cognition is through examining examples of student work (Fennema et al., 1996). Student work has also been the subject of research related to the benefits of portfolio assessments (Wolf, 1989). The purpose of such work is to form an alternative outcome measure to conventional testing. Although this is a different purpose from the present research—we seek to use student work as an implementation measure, not an outcome measure—one useful commonality is the technique of developing rubrics to score student work.

## Background: Scaling Up SimCalc

The Scaling Up SimCalc Project (Roschelle et al., in press) implemented two randomized controlled experiments (with one embedded quasi-experiment) designed to address the broad research question, "Can a wide variety of teachers use an integration of technology, curriculum, and professional development to increase student learning of complex and conceptually difficult mathematics?" There were two interventions, one for seventh grade and one for eighth grade. Each intervention integrated the representational technology SimCalc MathWorlds, curriculum workbooks, and teacher professional development organized around a 2-3 week replacement unit on rate, proportionality, and linear function. The replacement units incorporated the following hallmarks of the SimCalc approach to the mathematics of change and variation:

1. Anchoring students' efforts to make sense of conceptually rich mathematics in their experience of familiar motions, which are portrayed as computer animations.
2. Engaging students in activities to make and analyze graphs that control animations.
3. Introducing piecewise linear functions as models of everyday situations with changing rates.
4. Connecting students' mathematical understanding of rate and proportionality across key mathematical representations (algebraic expressions, tables, graphs) and familiar representations (narrative stories and animations of motion).
5. Structuring pedagogy around a cycle that asks students to make and compare their predictions.

The main effects of the treatment in the three main studies were positive, with student-level effect sizes of .63, .50 and .56; classrooms that used a SimCalc replacement unit had students who learned more advanced mathematics. This article focuses on data from the 48 treatment classrooms in the first experiment, which centered around seventh-grade students, teachers, and mathematical content.

SimCalc's developers have always espoused the view that students must be actively engaged with doing mathematics in order to learn; the software is a representational infrastructure that supports doing mathematics (Kaput, Hegedus, & Lesh, 2007). We confirmed this view, albeit weakly, in the larger quantitative data set by a finding that the number of days in the computer lab (a proxy for the level of student use of the software) correlated with overall learning gains. Case studies conducted within the Scaling Up SimCalc research provided further opportunities to examine this view within varying classroom implementations. Empson, Greenstein, Maldonado, and Roschelle (2009) examined differential student learning in three different classrooms, each with a different style of implementation. The analysis highlighted access to or blocking of learning resources as critical to students' ability to participate in the classroom as cognitively engaged learners. When a teacher significantly blocked students' autonomous use of SimCalc software and workbooks, by frequently interrupting the students' independent work, the classroom learning gains were lower. Likewise, Dunn (2009) found one low performing teacher spent much more time than the average teacher introducing and demonstrating ideas and much less time allowing students to work with the computer and workbooks. Analysis of student workbooks provides an opportunity to go beyond these case studies by examining connections between student completion of workbooks (and hence their level of engagement in doing the mathematics) with student learning.

SimCalc's developers have also always emphasized the activity of asking students to construct stories about graphs and motions (Kaput & Roschelle, 1998). These stories are seen as valuable because they evoke student engagement when they focus students on mathematically relevant details. For example, whether a students' story is about a race or a soccer game is irrelevant; whether a students' story describes speed and direction of motion is relevant. An emphasis on connecting linguistic and graphical forms of meaning is broadly consistent with the "Multimedia Principle" (Fletcher & Tobias, 2005). However, to date, there has been no empirical confirmation of this principle specific to SimCalc. Likewise, SimCalc developers believe that asking students to draw graphs and to explain motions are valuable activities. Analysis of student workbooks provides an opportunity to code these aspects of student work and to examine correlations to learning gains.

## SimCalc Workbooks

The seventh-grade curriculum, entitled *Managing the Soccer Team*, was designed to be used daily over a 2- to 3-week period to address central concepts of proportionality. Speed as rate is developed through a sequence of increasingly complicated simulation-based activities. The workbook is 59 pages long and has a total of 20 lessons. Lessons progress through representations—from graphs to tables to equations—aiming to teach students to translate among representations and to connect each concept to verbal descriptions of motion or other real-world contexts.

## Data Collection

Teachers were recruited for the Scaling Up SimCalc experiments from several regions of Texas. Recruitment and sampling in the overall experiment is described in Roschelle et al., in press. The seventh-grade treatment sample consisted of 796 students in 48 teachers' classes (for each teacher we collected data for one randomly selected class). The student ethnic breakdown was 48.5% White, 44.3% Hispanic, 4.2% African American, and 1.5% Asian. The mean campus-level percentage of students qualifying for free or reduced-price lunch was 54%, with a range of 1% to 94%, indicating a wide variation in campus poverty levels.

We asked all teachers participating in the Scaling up SimCalc experiment to return student workbooks to us and provided a prepaid express mail shipping box for them to do so. For the seventh-grade treatment teachers, we received boxes of workbooks from all 48 teachers, resulting in a collection of 765 workbooks (overall, 3.9% of the student workbooks were missing). Before analysis, each workbook was assigned the identification number that cross-referenced all other student-level data.

In addition to the workbooks, other data collected included: (1) unit pretests and posttests, (2) student and teacher demographics, (3) teacher mathematical knowledge, and (4) teacher daily implementation logs. Instrumentation and findings for these data are described extensively elsewhere (e.g., Roschelle et al., in press).

## Coding Protocol and Process

Based on a pilot workbook analysis, we designed a coding protocol that would be implemented by experienced math teachers. Teachers coded workbooks grouped by class so that they could observe patterns that may emerge within a class. There were three parts to the protocol.

1. For each workbook, the coder reviewed each of the 20 lessons and applied criteria to determine the level of completeness of the activities in the lesson (no attempt, low, medium, high). Coders did not make judgments about the quality of the work, but rather the presence of any attempted work.
2. For each workbook, the coder reviewed a subset of six key lessons to code the written work for correctness. The research team selected the six activities that seemed most central to the development of students' conceptual understanding. Four of these lessons entailed students' construction of stories from linear and/or piecewise linear graphs or construction of linear and/or piecewise linear graphs from stories. To capture the specificity of mathematical detail for each of the constructed stories and graphs, the coding protocol entailed a checklist of all possible correct mathematical details.
3. After coding a classroom for completeness and correctness, the coders made a holistic guess about the prior achievement level and learning gains in the classroom. The coders made written documentation of the indicators they used to make these judgments and recorded misconceptions that they observed.

We hired three master teachers who had not participated in the experiment to serve as coders during their school summer break. All were mathematics teachers, with one working at the high school level and two working at the middle school level. They all had little or no experience in conducting research.

The coders were trained on a set of 20 workbooks. They coded 5 workbooks together with one of the researchers and then 15 on their own. Interrater reliability on Parts 1 and 2 was sufficiently high for coders to then begin coding the full corpus. Within each class, two workbooks were randomly selected to be double-coded for Parts 1 and 2 to check for the maintenance of reliability.

## Analysis

We report our data analysis by each research question.

### 1. How long does it take to code workbooks?

We found it took the three teachers about 2 months working roughly half time to code 765 workbooks, including 3 days for training and 1 day for debriefing. The approximate time per workbook was 10–20 minutes, which translates to an average of about 4 hours per classroom, for classrooms implementing a 2–3 week long curriculum workbook. This was a relatively inexpensive task in the context of a multimillion dollar project.

### 2. Were we able to achieve adequate interrater reliability?

Interrater reliability was high. The index that we used was the intraclass correlation between each of the coder's ratings. Among the completeness and correctness ratings, there was a total of 26 ratings. The intraclass correlations for 20 of these ratings was over .90, 5 were between .80 and .90, and 1 was .66.

### 3. Could raters accurately detect indicators of high-achieving and low-achieving classrooms?

As shown in Figure 1, we found that coders were somewhat accurate in their holistic guesses about classrooms' average prior achievement levels [$t(33) = 2.5$, $p < .05$]; however they were not accurate in their holistic guesses about classrooms' average learning *gains*. We assume this is indicative of the characteristics of the student work, not these particular teachers' ability to detect learning gains.

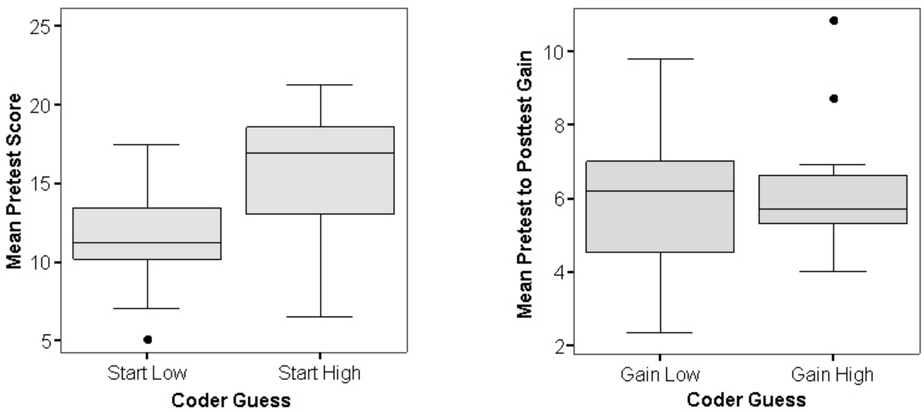



Figure 1. Actual classroom-level means distributed over coders' holistic guesses of achievement levels.

As summarized in Table 1, we debriefed the coders after completing this task to ascertain the qualities of student work that were salient to coders as they made these judgments. In addition, teachers were asked to collect misconceptions they observed in the workbooks. We found they detected a wide range of common misconceptions for this mathematical subject matter, including:

- Confusion between miles and mph—not paying attention to the fact that these are different.
- Distance interpreted as speed and speed interpreted as distance.
- Interpreting an object moving at a constant rate as increasing in speed.
- Confusion regarding which runner is the winner of a race based on reading a position graph.
- Confusion between dollars and decimals (e.g., "0.80 cents").
- In a position graph representing two races, the graphed line "on top" is more.
- Iconic interpretation of graphs as a picture (e.g., "the runner went up" describes a positively sloped position graph).
- In a piecewise linear function, two segments interchanged or confused on graph.
- Incorrect use of conventions (e.g., "13-50" as the ordered pair (13,50)).

Table 1. Qualities of student work that were salient to coders in judging student prior achievement and learning.

| Indicative of High Achievement | Indicative of Low Achievement |
|---|---|
| • Used pencil and erased when they made a mistake.<br>• Attention to correcting their work.<br>• Students were allowed to work on their own. The stories within a classroom were very different from each other, compared to classes where all the kids had the same stories verbatim.<br>• Indication of testing predictions with the software. Say that they saw their prediction was wrong.<br>• Quantified statements, rather than just saying "faster."<br>• Graphing was detailed. For every hour they would plot a point. Paid attention to precision, making sure they ended at the right point on the graph. Used straight edges.<br>• Stories included mathematical observations from the graph. Wrote a lot and were very specific (not simply a creative story).<br>• Evidence of teacher feedback and modeling. | • Sloppy handwriting, incomplete answers, poor spelling, reverse letters, graphing without a straight-edge, wrote in pen, did not correct answers.<br>• Lack of detail, not paying attention to answering questions completely.<br>• A lot of absences—large numbers of pages with missing work.<br>• Struggled with math vocabulary and writing skills. Could not accurately describe motion because they had poor word choice. "The van caught up" instead of "the bus slows down" was very important.<br>• Stories tended to be more creative but were missing a lot of mathematical detail.<br>• In some classrooms, all students had the same incorrect answers. This suggested a teacher error or misconception. |

## 4. Does completeness of student workbooks predict learning gains?

Overall, we found that the mean completion of student workbooks was 75% (SD = 17%), which suggests that most teachers really did implement the SimCalc intervention. As Figure 2 shows, there was considerable variation in workbook completion, both within and across classrooms. The figure shows classrooms in rank order of the classroom median for student completion of workbooks. About two-thirds of the classrooms completed 75% or more of the workbooks. Of the remaining third, three classrooms completed 50% or less of the workbook.

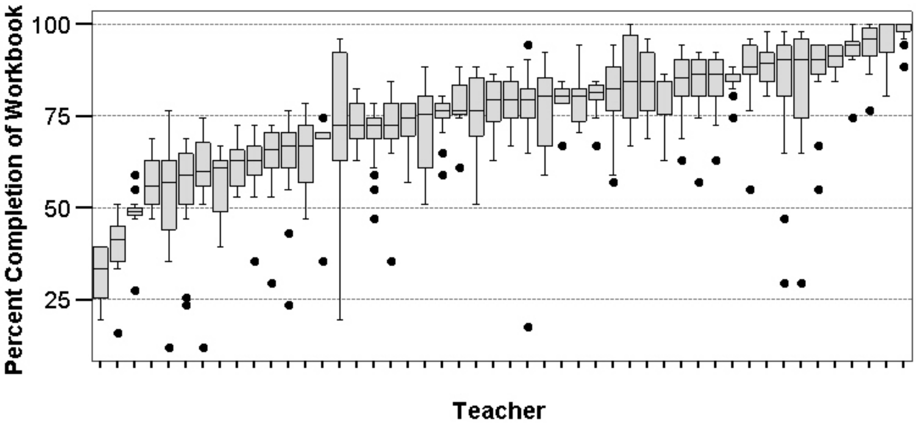



Figure 2. Distributions across classrooms of completion of workbook (teachers ordered by class median).

We then tested whether workbook completion predicted pretest to posttest learning gains, using a hierarchical learning model (HLM) to correct for the nesting of students within classrooms. We examined the "$M_2$" subscale of our assessment, which focuses on more advanced mathematics. We chose to focus on this scale because classrooms that used SimCalc had particularly large gains on this scale; hence, it is sensible to look at this scale when considering the value added by the SimCalc intervention. As Figure 3 shows, there was an overall correlation between student completion of workbooks and mean classroom learning gains [$z = 4.3$, $p < .001$]. Using HLM, we determined that on the 21-point $M_2$ scale, students on the average gained 1 $M_2$ point for every 13.3% of the workbook that was completed.

We then examined completion of each of the 20 lessons individually. We found that the completion rate was uniformly high for the first half of the workbook (with the exception of a section intended for homework); declines in workbook completion occurred in some classrooms in the second half of the workbook. We found a high alpha (0.83) among completion rates for all activities, suggesting that whether a student

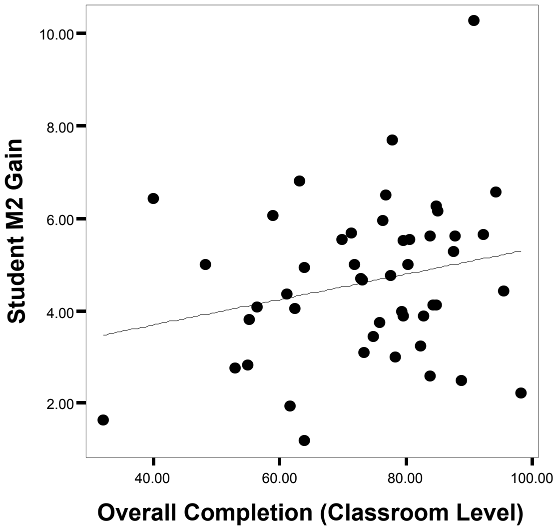



Figure 3. Workbook completion and student learning gains.

completes one activity is a good indicator of whether they will complete other activities. With respect to learning gains, we found correlations between completion of most activities in the second half of the curriculum and learning gains. This may be due to the fact that the second half of the workbook covered the material most relevant to the $M_2$ scale and/or had the most variation in completion.

We then examined factors that might predict completeness of the workbooks. The best predictor of workbook completeness was students' pretest score. The pretest had 30 items. For every 10 additional items correct, students on average completed an additional 6% of the workbook [$z = 5.92$, $p < .001$]. The other statistically significant predictor was gender. On average, girls completed 1.8% more of the workbook than boys did. The other variables that were not significant included: geographic location, student ethnicity, school poverty level, teachers' years-of-experience, teachers' level of mathematical knowledge, how often class was conducted in the computer lab, and the number of days spent implementing the unit. Clearly workbook completion is not a proxy for computer use.

We also used the number of days the unit was implemented in conjunction with the completeness to examine the *pace* with which the class did the unit. Overall, workbook completion was independent of the time teachers' spent doing the unit. However, we did find that classrooms with a higher mean pretest score completed the workbooks faster [$r(48) = .42$, $p < .01$]. Teachers may be purposefully taking more or less time with the workbooks depending on the incoming level of their students, which suggests that most teachers managed implementation to substantially complete the workbooks rather than to occupy a fixed amount of time.

## 5. Does correctness of details in workbook answers predict learning gains?

Figure 4 shows examples of two students' written stories in response to the graph depicted in the workbook. As we found throughout the workbooks, students varied in how much mathematical detail they provided and whether the detail was correct. In the top story, the student provides little detail about speed and describes the wrong vehicle as stopped. In the bottom story (which is still relatively incomplete compared to other stories we saw), the student describes the constant speed of the bus until it stopped. (The scores "-30" and "-10" were written by the teacher and not part of our coding scheme.) We see that the student in the top story has also made notes on the task: "explain lines using speed, location, time; explain motion." One lesson we learned from reviewing workbooks is that the task may not have been clear to most students and teachers; this teacher may have given the students more information on what they were supposed to do. In a revision of the materials, we would consider more specifically prompting students to focus on mathematical detail.

Overall, we found that the level of mathematical detail and correctness in all six activities predicted student learning, either significantly or marginally significantly. Also, as with the completion rates, we found high intercorrelations between correctness for each lesson (alpha = 0.70). Hence, our ability to argue that doing any specific learning activity is important to learning is limited. However, on a whole this finding does suggest that the activities that focus on drawing graphs and writing stories are valuable for student learning.

As was the case for completeness, student pretest scores predicted correctness in the workbook. No other variable predicted correctness, including student gender, geographic location, student ethnicity, school poverty level, teachers' years-of-experience, teachers' level of mathematical knowledge, how often class was conducted in the computer lab, and the number of days spent implementing the unit.

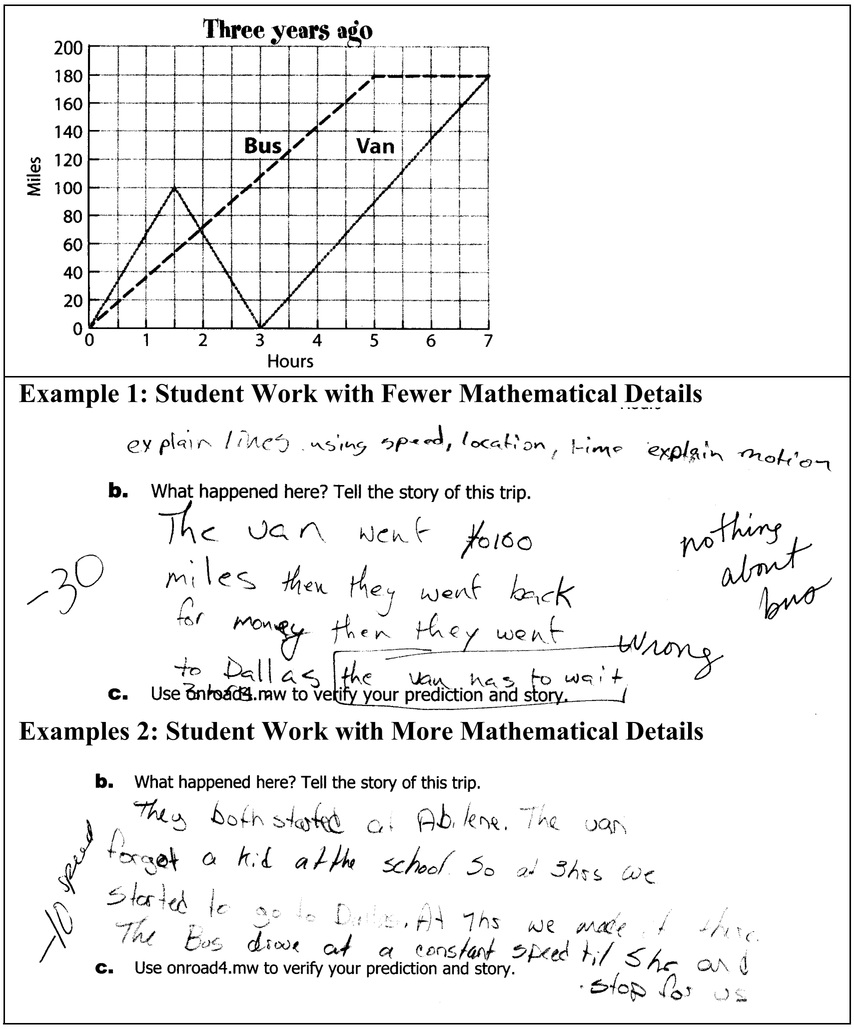



Figure 4: Two contrasting student responses to a story writing task.

## Discussion

Our analysis of the first two research questions suggests that analysis of student workbooks is a viable technique for measuring implementation of a Learning Sciences–based innovation at scale. Workbook coding is reasonably cost effective and time efficient, especially relative to methods requiring field observations or video recordings of classrooms. Furthermore, we were able to achieve suitable interrater reliability. Workbook coding has the advantage that it makes it possible to examine the work of the majority of the students in a classroom, whereas field observations or video recordings often have to be focused on just a few students.

The third research question looked at the connections between coders' holistic judgments and students' actual knowledge and learning. We found that coders could detect with some accuracy which classrooms had lower mean pretest scores. However, they did not tend to detect which classrooms had lower or higher mean learning gains. This suggests that coding specific activities for completeness and correctness provides important analytical information that cannot necessarily be obtained through teachers' holistic judgments.

We did find that completeness of student workbooks predicted student learning. This is important because it suggests that having students do the mathematical work is important to student learning, a core tenet of SimCalc program developers. Although this may seem obvious, it is not obvious to all teachers who implement SimCalc—some prefer to focus on teacher demonstrations using the SimCalc software. Indeed, case studies provide evidence of some classrooms in which teachers blocked student independent work; this work confirms the conjecture arising from those case studies that the workbooks are a valuable learning resource and that low use of the workbooks was a factor in lower student learning gains. A corollary finding was that pretest scores predicted completeness. This is an indication of an advantage that students with higher prior knowledge have in learning the content of this unit. This suggests the necessity for additional supports for students who start the unit with lower prior knowledge, a point that could be addressed in teacher professional development to improve implementation.

We also found that the level of detail in student work and the correctness of the detail predicted student learning. We particularly focused on details in students' construction of graphs and stories about graphs. This is the first empirical confirmation that these activities, long part of the SimCalc approach, have an impact on learning.

Of course, because the workbook is designed for and used in a variety of ways as part of the many resources in the classroom (i.e., not as an assessment instrument), inferences about the work must be made with caution. For example, it is possible that students can be actively engaged in doing simulations with the software or having mathematical conversations with the teacher or peers yet not write extensively in their workbooks. However, while these limitations do exist, we believe the methodology does provide an important window into students' cognitive engagement across a large-scale study that would otherwise be unavailable.

Overall, we would recommend that Learning Scientists who are scaling up their curricular designs consider designing workbooks and other student work artifacts that can be collected and analyzed for evidence of the quality of classroom implementations. Relative to field observation or video recording, analysis of student work is relatively cost effective, while preserving the opportunity for detailed insight into how teachers and students are using materials. This form of data collection may be particularly useful in examining the relationship of constructive activities, such as drawing graphs and writing stories, to students' learning gains.

## Acknowledgements

This material is based in part upon work supported by the National Science Foundation under Grant Number 04-37861. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also acknowledge and deeply appreciate our three master teachers for all of their hard work in doing this coding.

## References

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. *Innovations in learning: New environments for education*, 289-325.

Clement, J. (1989). The concept of variation and misconceptions in Cartesian graphing. *Focus on Learning Problems in Mathematics, 11(*1-2), 77-87.

Cohen, D. K., & Hill, H. (2001). *Learning policy: When state education reform works.* New Haven, CT: Yale University Press.

Dunn, M. B. (2009). *Investigating variation in teaching with technology-rich interventions: What matters in training and teaching at scale?* Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.

Empson, S. B., Greenstein, S., Maldonado, L., & Roschelle J. (2009). Scaling up innovative mathematics in the middle grades: Case studies of "good enough" enactments. Manuscript submitted for publication.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403-434.

Fletcher, J. D., & Tobias, S. (2005). The multimedia principle. *The Cambridge handbook of multimedia learning*, 117–133.

Kaput, J., Hegedus, S., & Lesh, R. (2007). Technology becoming infrastructural in mathematics education. In R. Lesh, E. Hamilton & J. Kaput (Eds.), *Foundations for the future in mathematics education* (pp. 173-192). Mahwah, NJ: Lawrence Erlbaum Associates.

Kaput, J., & Roschelle, J. (1998). The mathematics of change and variation from a millennial perspective: New content, new context. In C. Hoyles, C. Morgan & G. Woodhouse (Eds.), *Rethinking the mathematics curriculum* (pp. ).. London, UK: Falmer Press.

Newman, F. M., Lopez, G., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago schools: A baseline report*. Chicago IL: Consortium on Chicago School Research.

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J. & Gallagher, L. (in press). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal.*

Roschelle, J., Tatar, D., Shechtman, N., & Knudsen, J. (2008). The role of scaling up research in designing for and evaluating robustness. *Educational Studies in Mathematics, 68*(2), 149-170.

Wolf, D. P. (1989). Portfolio assessment: Sampling student work. *Educational Leadership, 46*(7), 35-39.

# Motivation to Transfer Revisited

Andreas Gegenfurtner, Marja Vauras, Centre for Learning Research, University of Turku, Finland
Email: angege@utu.fi, vauras@utu.fi
Hans Gruber, Institute of Educational Sciences, University of Regensburg, Germany, Email:
hans.gruber@paedagogik.uni-regensburg.de
Dagmar Festner, f-bb Research Institute for Vocational Education and Training, Nuremberg, Germany, Email:
festner.dagmar@f-bb.de

**Abstract:** Should the construct 'motivation to transfer' used in human resource development and management research be also used in learning research? The current study revisited motivation to transfer on a sample of 128 participants of occupational health training. Confirmatory factor analysis and partial least squares-based path modeling were used to test the hypothesized dimensions and relationships among variables including social and affective cues on training transfer. Based on a combination of the theory of planned behavior, expectancy theory, and self-determination theory, we validated three dimensions of transfer motivation: autonomous motivation to transfer, controlled motivation to transfer, and intentions to transfer. Results indicate that autonomous motivation was affected by attitudes toward training content and utility reactions; controlled motivation was affected by utility reactions, supervisory support, and social norms. Intentions to transfer mediated the effects of autonomous motivation on transfer three months after training. Implications of a multidisciplinary perspective combining learning sciences and human resource development are discussed.

## The Noe Model Revisited

In 2006, in a discussion on how motivation influences transfer, Pugh and Bergin (p. 155) recommended adopting the construct motivation to transfer from the human resource development (HRD) literature to the learning sciences; their recommendation was published in a highly influential journal in the field, *Educational Psychologist*. Twenty years earlier, Raymond Noe coined the term 'motivation to transfer' that Pugh and Bergin were referring to as "the trainees' desire to use the knowledge and skills mastered in the training program on the job" (Noe, 1986, p. 743). His work was published in a highly influential journal in the field of management and HRD, *Academy of Management Review*. It is argued here that it would be fatal if learning scientists followed Pugh and Bergin's advice to adopt the construct from HRD research to learning research. This is for five reasons. First, motivation to transfer currently is a one-dimensional construct that is too coarse to adequately reflect motivational states and traits in the transfer process. Second, motivation to transfer currently is not, and was not, adequately grounded in motivation theories. Third, there is poor empirical evidence in the HRD and management literature on how motivation to transfer affects transfer of learning. Fourth, motivation to transfer currently is a static construct, only measured immediately following an intervention. Finally, Noe (1986) said little in his original conceptualization about the social origin of transfer motivation; this reflected the zeitgeist of the 1980s epoch. Learning researchers now can be thoughtful in adopting a more balanced view.

It is argued further that motivation to transfer does indeed play a central role in the transfer process; there is enough theoretical (and common sense) backdrop to expect that learners who are motivated to apply training content from training to work, or from the classroom to out-of-classroom use, are more likely to be successful than learners who are not motivated. Yet, in its current form, motivation to transfer seems to be a construct powerless to account for intentional transfer processes. Transfer, from a human resource perspective, is defined as the successful application of newly trained knowledge, skills, and attitudes to the workplace (Noe, 1986; Burke & Hutchins, 2007). Hence, we argue that before following Pugh's and Bergin's recommendation, Noe's model of motivation to transfer must be revisited in several significant respects: dimensionality, theoretical grounding, empirical evidence, measurement time, and social nature of motivation.

On revisiting dimensionality, motivation to transfer has been measured as a one-dimensional construct in empirical training research over the past twenty-five years or so (Noe & Schmitt, 1986; Velada, Caetano, Bates, & Holton, 2009). In our recent review on motivation to transfer (Gegenfurtner, Veermans, Festner, & Gruber, 2009), we found that all 31 studies on transfer motivation reported in international peer-reviewed journals from 1986 to now used a one-dimensional scale to assess the construct, ranging from one item to eleven items. This is a paradox since motivation researchers have explored numerous dimensions of motivational processes in human actions. To name just a few, we know of intrinsic and extrinsic motivation; conscious goal intentions and unconscious implementation intentions; expectancies, instrumentalities, and valences; various types of motivational regulation and mindsets; and we differentiate motivation, volition, and emotion. All contribute to our understanding of the many facets and colors that motivation has. Motivation to transfer potentially goes in concert with all of the dimensions just mentioned. It can be argued that motivation to transfer has remained a one-dimensional construct to date because it lacks a clear grounding in motivation theories.

On revisiting theoretical grounding, Noe (1986; Noe & Schmitt, 1986) made no explicit statement as to which theoretical framework the idea of motivation to transfer is based; although there are references to Bandura, Latham, and Vroom in their original papers, the associated motivation theories refer not to transfer motivation but to other aspects of the papers. Hence, motivation to transfer lacked from the start, and still lacks, a solid foundation in motivation theories. In an attempt to provide a first step toward resolving this situation, we conducted two studies in which the theory of planned behavior (Gegenfurtner & Gruber, 2009) and expectancy and self-determination theory (Gegenfurtner, Festner, Gallenberger, Lehtinen, & Gruber, 2009) were tested as a basis for transfer motivation scales. In short, these studies indicate that motivation to transfer has several dimensions: autonomous motivation to transfer, controlled motivation to transfer, and intentions to transfer. These are the first steps, and more efforts are needed to validate the construct's theoretical grounding for human resource development beyond.

On revisiting empirical evidence, it is a likely explanation that the one-dimensionality and the lack of theoretical grounding account for the current state of poor evidence on the intention-behavior relationship in training transfer. Correlation coefficients ranging from 0.04 (Tziner, Haccoun, & Kadish, 1991) to 0.63 (Machin & Fogarty, 1997) suggest that the relation between motivation to transfer and transfer of training needs further elaboration. Again, we believe that a starting point for this elaboration is a reconceptualization of Noe's model of transfer motivation. This may help to form new research practices with respect to the multidimensional nature of transfer motivation, its groundwork in valid motivation theories, and its measurement time.

On revisiting measurement time, our review of transfer motivation (Gegenfurtner et al., 2009b) indicated that, in 30 of 31 cases, the construct was assessed at the immediate end of training. Only Leitl and Zempel-Dohmen (2006) measured transfer motivation at the immediate end *and* at a later point in time, three months after the intervention. We believe that motivation to transfer in particular, like motivation for action in general, dynamically changes over time. Hence, assessment of transfer motivation at a time when the learner has had time to explore opportunities to use training in out-of-classroom applications may provide a very different picture than the assessment of transfer motivation at a time when the learner is just about to leave the classroom. Yet, empirical examinations at a different time than at the immediate end of training are almost non-existent.

On revisiting the social nature of motivation, there is large consensus that motivation is intrinsically social in nature (Hickey, Moore, & Pellegrino, 2001; Järvelä, Volet, & Järvenoja, 2010; Vauras, Salonen, & Kinnunen, 2008). We argue that this holds equally for transfer motivation. Especially when trainees spent some time at the workplace and could test the training content about its usefulness, social cues might affect formation and persistence of motivational states to transfer training. Arguably, the generation of more stable motivational traits is also influenced by factors such as social norms, feelings of relatedness, or control beliefs that are based on environmental (working) conditions emanated in social interaction. Noe (1986) constructed his model of transfer motivation at a time in which the social nature of motivation was not on the agenda yet. Reconceptualizing the Noe model can account for this development in motivation research.

## Present Study

The present study had a dual goal. First, it aimed to test the multidimensionality of motivation to transfer. Multidimensionality was achieved by using a combined motivation theory approach. Specifically, there was a combination of a validated framework from educational psychology, self-determination theory (Baard, Deci, & Ryan, 2004), and a validated framework from management research, expectancy theory (Vroom, 2005), to conceptualize autonomous and controlled motivation to transfer. Expectancy theory as a cognitive-choice approach and self-determination theory as a need-motive-value approach can complement each other in predicting and explaining human performance in the workplace (Kanfer, 1990). Thus, instrumentality and valence items reflecting externally prompted reasons to transfer were used to assess controlled motivation. Conversely, instrumentality and valence items reflecting internally regulated behaviour were used to assess autonomous motivation. Autonomous motivation to transfer is defined here as an internalized desire to transfer learning that is initiated and governed by the self (i.e., regulated by identification or by integration with one's values), and controlled motivation as a desire to transfer learning that is not initiated and governed by the self (i.e., regulated by external rewards or sanctions). In addition to autonomous and controlled motivation as measures of motivational traits, intentions to transfer are also included as measures of motivational states. This is because contrary to motivational traits, intentions represent a more activated, situation-specific motivational state. We used Ajzen's (1991) theory of planned behaviour, a well-validated framework from social psychology, to conceptualize intentions to transfer.

The second goal was to test a hypothesized path model, shown in Figure 1. The model explored a motivational sequence. Effects of autonomous and controlled motivation on training transfer are mediated by the more situation-specific state of transfer intentions; at the same time, situational but distal social (relatedness, support, control, and norms) and affective (attitudes and utility reactions) cues on transfer intentions are mediated by more stable motivational traits, i.e., autonomous and controlled motivation to transfer. It can be argued that motivational traits act like filters. For example, supervisory support may or may not lead to activated intentions, depending on the motivational trait. The model has a range of hypothesized relationships; these are based on theory or past empirical evidence. Specifically, the hypothesis, based on self-determination theory, predicted that relatedness (RE) at work would foster internalization of external regulations and would thus be positively related to autonomous motivation to transfer (Baard et al., 2004; Gegenfurtner et al., 2009a). Supervisory support (SU) was shown to have mixed empirical results on motivation, depending on whether support was perceived to be instrumental or not; we thus expected relations to both autonomous and controlled motivation. Based on the theory of planned behavior (Ajzen, 1991), it was hypothesized that perceived behavioral control (PBC) promoted trainees' feelings of autonomy and would thus be more important for autonomous motives; at the same time, social norms (NO) were hypothesized to have stronger effects for those trainees with controlled motivation. Past research showed that attitudes toward training content (AT) were related to both autonomous and controlled motivation (Gegenfurtner et al., 2009a). Utility reactions (UT) were hypothesized to affect self-determined feelings of autonomy; however, due to their instrumental nature, it was speculated that they could also affect controlled motivation.

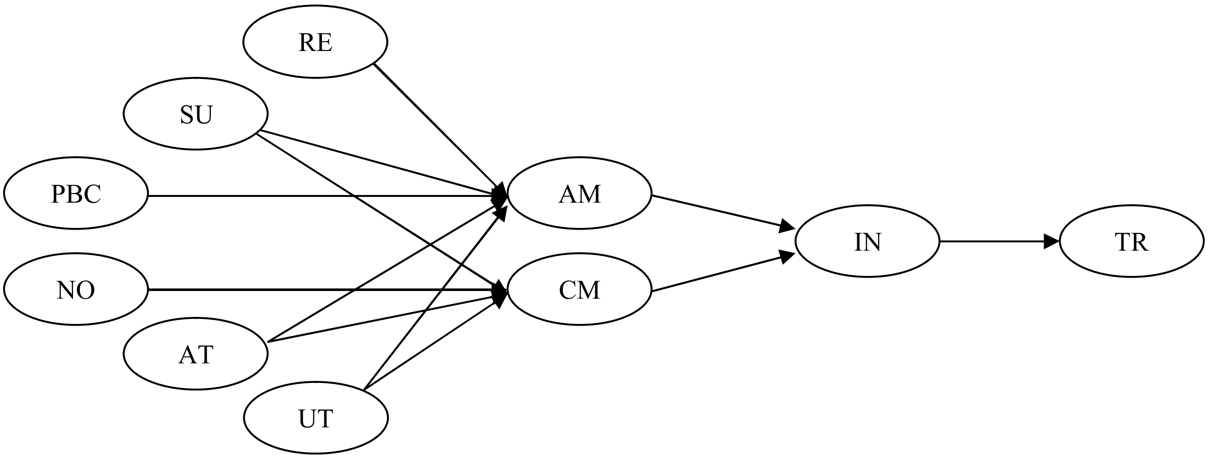



Figure 1. Hypothesized path model.

# Method

## Participants and Procedure

The participants in the study were 496 employees who attended one of the 23 training programs on occupational health and safety following German statutory accident insurance regulations. Three months after the training, trainees received a paper-and-pencil questionnaire sent to their workplace. The questionnaire collected self-report data that, despite the known problems (e.g., leniency, self-serving bias), seemed adequate to use because a major interest was trainees' self-perceived attitudinal and motivational states. One hundred thirty-one trainees (26.4%) completed and returned the questionnaire. Deletion of three multivariate outlying cases ultimately yielded a final sample size of 128 trainees. The majority of the participants was between 41 and 50 years old (43.1%) and has worked with their current employer for up to five years (25.0%); the average organization size was between 100 and 199 employees (68.8%). There was no statistically significant difference between respondents and non-respondents. Participation in the study was voluntary, with informed consent forms signed by the trainees. Anonymity and confidentiality were guaranteed for all responses.

## Instruments

Unless otherwise indicated, a 5-point response scale was used for all items, with 1 = *strongly does not agree*, 2 = *does not agree*, 3 = *partly agrees*, 4 = *agrees*, and 5 = *strongly agrees*. All scales were pilot-tested; minor revisions in expressions and the ordering of items were reflected in the final version of the instrument. Table 1 presents internal consistency values of each scale.

### Demographic and Organizational Membership Characteristics

Demographic and organizational membership characteristics were measured to establish nomological validity. One of each item was used to assess organizational tenure ('How long have you worked with your current employer?'; 1= up to 5 years, 2=6-10, 3=11-15, 4=16-20, 5=21 years or more), organization size ('How many people are employed in your company?'; 1=1-9 employees, 2=10-19, 3=20-49, 4=50-99, 5=100-199, 6=200-249, 7=250-349, 8=350-549, 9=550-999, 10=1000-1999, 11=2000 employees or more), and age (1=up to 30, 2=31-40, 3=41-50, 4=51-60, 5=more than 60 years old).

## Independent Variables: Social and Affective Cues

This study used social and affective cues as independent variables. First, social cues were measured with scales on relatedness, supervisory support, perceived behavioral control, and social norms. To assess relatedness, five items were used based on the Basic Psychological Needs Scale at Work (Baard et al., 2004; Gegenfurtner et al., 2009a); the items measured the extent to which trainees felt connected and respected in their organizations (e.g., 'People at work care about me'). To assess supervisory support, five items were used that measured support associated with goal-setting and feedback (e.g., 'I try to achieve preset goals by applying what I have learned in training on the job'). To assess perceived behavioral control, we included five items that measured participants' perceived capability of and confidence in transferring the training to their workplace (e.g., 'I feel capable to apply the training content at work'). To assess social norms, five items were used that reflected the degree to which participants perceived social normative pressure from referents important to the trainees to perform or not to perform a certain behavior (e.g., 'My boss sets a high value on applying the training content at work').

Second, affective cues were measured with scales on utility reactions and attitudes toward training content. To assess utility reactions, items developed by Tai (2006) were used that measured the affective experience on whether the training has been useful during the last three months (e.g., 'Today, I judge the usefulness of the training as: 1=very low, 2=low, 3=average, 4=high, 5=very high'). To assess attitudes toward training content, five items were used to measure cognitive and emotional/affective aspects of attitudes towards occupational health and safety (Ajzen, 2001; Eagly & Chaiken, 1993; Gegenfurtner et al., 2009a). Sample items were as follows: for cognition, 'I easily come up with at least five reasons for complying with safety and health regulations', and for emotion, 'I feel responsible for health and safety in my working area'.

## Mediating Variables: Motivation to Transfer

Motivation to transfer was measured with three distinct scales specifying autonomous motivation to transfer, controlled motivation to transfer, and intention to transfer (Gegenfurtner et al., 2009a; Gegenfurtner & Gruber, 2009). For autonomous motivation, two pairs of items were included to measure instrumentality and valence of autonomous motives for training transfer. Sample items were as follows: for instrumentality, 'While applying training at work, I can learn a lot', and for valence, 'This learning is important to me'. For controlled motivation, two pairs of items were included to measure instrumentality and valence of controlled motives for training transfer. Sample items were as follows: for instrumentality, 'Successful application of the training content will probably result in a materialistic reward, such as a financial bonus', and for valence, 'This reward is important to me'. A five-item scale was used to measure intention to transfer (e.g., 'I have tried intentionally to apply the training content to my workplace, regardless how well it actually worked').

## Dependent Variable: Transfer of Training

To assess transfer of training, five items described in Festner and Gruber (2008) were used to measure self-perceived change in knowledge, skills, and attitudes toward the training content, occupational health and safety (e.g., 'Today, I engage more in health and safety than I did before the training').

# Data Analyses

A two-stage procedure was adopted for data analysis. First, the factorial validity of the revisited construct of motivation to transfer was assessed. This was done using confirmatory factor analysis (CFA). Given an appropriate level of communality, CFA was considered meaningful (MacCallum, Widaman, Zhang, & Hong, 1999). Based on the EQS 6.1 (Bentler, 2005) software, three first-order CFA models designed to test the multidimensionality of the theoretical construct were examined. The first model was a three-factor model composed of autonomous motivation to transfer, controlled motivation to transfer, and intention to transfer. The second model was a two-factor model in which autonomous and controlled motivation were merged as if representing one factor. Finally, the third model was a one-factor model that forced autonomous motivation to transfer, controlled motivation to transfer, and intention to transfer into one factor (which is the current research practice). Data were screened to test for multivariate outliers, normality and multi-collinearity (Kline, 2005). The direct maximum likelihood approach was used as a missing data specification procedure and robust methods as normality estimator corrections. Assessment of the model fit was based on four criteria reflecting statistical and theoretical considerations. The criteria were as follows: (1) the Yuan-Bentler scaled $\chi^2$ test statistic, (2) the comparative fit index (CFI), (3) the standardized root-mean square residual (SRMR), and (4) the root-mean square error of approximation (RMSEA), with its 90% confidence interval (CI). This rationale was based upon literature recommendations (Bentler, 2005; Kline, 2005). For cut-off criteria, guidelines were followed for CFI>0.95, SRMR<0.09, and the RMSEA<0.06 (Hu & Bentler, 1999) to indicate appropriate goodness-of-fit. Validation of the three-dimensional construct seemed important to test the extent to which the measured variables actually represent the theoretical dimensions.

Once the factorial validity of motivation to transfer was established, its mediating position was assessed in our hypothesized model. This was done using path analysis following a partial least squares (PLS) approach. Contrary to other estimation techniques like multiple regression or structural equation modeling, PLS can be applied to a non-normally distributed data set collected with a small sample (Chin & Newstead, 1999). Based on the SmartPLS 2.0 (Ringle, Wende, & Will, 2005) software, the relationships among the variables were assessed using the path weighting scheme algorithm. Importantly, PLS is an approach for predicting relationships in a model, not for assessing overall model fit. However, Table 1 reports three reliability indices to indicate appropriate psychometrics properties of the measurement models. For cutoff-criteria, guidelines were followed for Cronbach's alpha>0.70, the average variance extracted (AVE)>0.50, and the composite scale reliability (CSR)>0.60 (Hair, Black, Babin, Anderson, & Tatham, 2006). Mediation was analyzed following the recommendations by MacKinnon, Fairchild, and Fritz (2007).

## Results

Screening of the data revealed no multicollinearity but multivariate nonnormality (Yuan, Lambert, and Fouladi's normalized estimate=45.88). Three multivariate outlying cases with a substantial different contribution to normalized multivariate kurtosis were subsequently deleted from all analyses. Table 1 presents psychometric properties and correlation coefficients of the measures. The next two sections describe test results for multidimensionality of motivation to transfer and for the hypothesized path model.

Table 1: Psychometric Properties and Correlation Coefficients of All Measures

| | M | SD | α | AVE | CSR | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | 2.67 | 0.88 | - | - | - | - | .12 | .02 | .03 | .06 | .03 | .02 | .01 | .02 | .06 | .02 | .01 | .05 |
| OT | 5.45 | 2.97 | - | - | - | .34 | - | .01 | .08 | .02 | .03 | .01 | .08 | .01 | .05 | .01 | .06 | .01 |
| OS | 2.71 | 1.39 | - | - | - | .04 | .11 | - | .03 | .04 | .03 | .02 | .01 | .01 | .01 | .01 | .01 | .05 |
| RE | 4.03 | .79 | .84 | .60 | .88 | .18 | .09 | .18 | - | .27 | .34 | .35 | .26 | .18 | .13 | .10 | .19 | .04 |
| SU | 3.69 | .97 | .71 | .69 | .76 | .08 | .04 | .06 | .52 | - | .25 | .42 | .32 | .34 | .30 | .34 | .34 | .19 |
| PBC | 4.10 | 1.01 | .84 | .78 | .82 | .17 | .05 | .16 | .58 | .59 | - | .41 | .41 | .41 | .31 | .20 | .40 | .09 |
| NO | 2.41 | 1.23 | .90 | .71 | .93 | .15 | .11 | .13 | .59 | .65 | .64 | - | .32 | .36 | .24 | .32 | .20 | .10 |
| AT | 3.60 | .85 | .73 | .68 | .82 | .12 | .09 | .11 | .51 | .57 | .64 | .57 | - | .31 | .32 | .20 | .35 | .17 |
| UT | 4.14 | .75 | .82 | .64 | .88 | .14 | .11 | .12 | .42 | .58 | .64 | .60 | .56 | - | .40 | .28 | .38 | .36 |
| AM | 4.12 | .84 | .83 | .66 | .88 | .08 | .07 | .10 | .36 | .55 | .56 | .49 | .57 | .63 | - | .35 | .37 | .40 |
| CM | 2.68 | 1.16 | .74 | .66 | .77 | -.04 | .11 | -.11 | .31 | .58 | .45 | .56 | .45 | .53 | .59 | - | .19 | .18 |
| IN | 4.02 | .95 | .73 | .68 | .82 | .11 | .08 | .12 | .44 | .58 | .63 | .46 | .59 | .62 | .61 | .44 | - | .23 |
| TR | 3.82 | 1.02 | .87 | .76 | .91 | -.10 | .12 | -.22 | .19 | .44 | .30 | .31 | .41 | .60 | .63 | .43 | .48 | - |

*Note.* Values below the diagonal are correlation estimates. Values above the diagonal are squared correlation estimates. OT=organizational tenure, OS=organization size, RE=relatedness, SU=supervisory support, PBC=perceived behavioral control, NO=social norms, AT=attitudes toward training content, UT=utility reactions, AM=autonomous motivation to transfer, CM=controlled motivation to transfer, IN=intention to transfer, TR=transfer of training $p<.05$ for $r>14$.

### Testing for Multidimensionality

The model to be tested a priori postulates that motivation to transfer three months after training is a three-factorial structure composed of autonomous motivation to transfer, controlled motivation to transfer, and intention to transfer. The three-factor model was tested first; then it was compared to the two-factor model and the one-factor model. Fit statistics relative to these models are presented in Table 2. In reviewing the CFI, SRMR, RMSEA, and $\chi 2$ test statistic values in Table 2, it is evident that the hypothesized three-factor model represented the best fit to the data.

Table 2: Confirmatory Factor Analytic Model Fit Statistics.

| Model | $\chi^2$ | *df* | CFI | SRMR | RMSEA (with 95% confidence interval) |
|---|---|---|---|---|---|
| Three-factor model | 112.24 | 62 | .97 | .08 | .05 (.04 - .08) |
| Two-factor model | 148.45 | 64 | .85 | .10 | .09 (.07 - .12) |
| One-factor model | 176.59 | 65 | .76 | .10 | .11 (.09 - .13) |

After CFA, construct validity of the three-factor solution was tested by assessing convergent, discriminant, nomological, and face validity. We followed the guidelines from Hair et al. (2006). First, concerning convergent validity, the variance-extracted measures (AVE) of AM, CM, and IN exceeded the 50% level, and the reliability estimates were larger than .70. Although two factor loadings were below .50, they did not appear to significantly harm the model fit or internal consistency; hence, convergent validity was established. Second, a construct is divergently valid if the variance-extracted estimates for each factor are larger than the squared interconstruct correlations associated with this factor (Hair et al., 2006). Table 1 reports AVE and squared intercorrelations. All variance-extracted estimates are larger than the corresponding squared intercorrelation estimates; hence, divergent validity of AM, CM, and IN was established. Third, concerning nomological validity, Hair and colleagues (2006) recommend comparing the constructs to other variables not included in the model, which, in this study, were demographic and organizational membership characteristics. The three factors were not significantly affected by trainee age, organizational tenure, or organization size; hence, nomological validity was established. Finally, face validity of the three dimensions was established based on discussions among the authors AG and DF about the content of the corresponding items. In summary, both the established construct validity and the acceptable model fit indicated good conditions to further test the hypothesized relationships of the three-factor solution in the PLS path model.

### Testing the Hypothesized Path Model

Table 3 presents the parameter estimates of the hypothesized path model. Mediation analysis (MacKinnon et al., 2007) indicates that intention to transfer mediated the effect of autonomous motivation on training transfer; autonomous motivation had a stronger effect on intentions than controlled motivation. Supervisory support, social norms, and utility reactions significantly affected controlled motivation while attitudes toward training content and utility reactions affected autonomous motivation. It is interesting to note that supervisory support seems to be effective for controlled motivation, but not for autonomous motivation. Despite the hypotheses, the paths from relatedness, support, and perceived behavioral control to autonomous motivation were non-significant, as was the path from controlled motivation to transfer to intention to transfer. This indicates that autonomous motives lead to more activated intentions than controlled motives three months after training.

Table 3: Confirmatory Factor Analytic Model Fit Statistics.

| Effect of | On | Path coefficient | *t*-value |
|---|---|---|---|
| Relatedness | Autonomous motivation | -.04 | 0.06 |
| Supervisory support | Autonomous motivation | .12 | 1.31 |
| Supervisory support | Controlled motivation | .29 | 3.04* |
| Perceived behavioral control | Autonomous motivation | .01 | 0.02 |
| Social norms | Controlled motivation | .24 | 2.11* |
| Attitudes toward training content | Autonomous motivation | .22 | 2.64* |
| Attitudes toward training content | Controlled motivation | .04 | 0.35 |
| Utility reactions | Autonomous motivation | .55 | 7.07* |
| Utility reactions | Controlled motivation | .21 | 2.27* |
| Autonomous motivation | Intention to transfer | .54 | 6.08* |
| Controlled motivation | Intention to transfer | .12 | 1.39 |
| Intention to transfer | Transfer of training | .48 | 6.39* |

*Note*. * $t > 1.96 = p < .05$

## Discussion and Conclusion

The theme of ICLS2010 invited an exploration of disciplinary perspectives and how multidisciplinary approaches can advance the learning sciences. This study explored research practices in the HRD and management disciplines with respect to motivation and transfer Revisiting the Noe model, it was argued that the construct ‘motivation to transfer’ needs to be reconceptualized in several significant respects before learning researchers should follow Pugh’s and Bergin’s (2006) recommendation to adopt motivation to transfer in contexts beyond HRD. This study used a multidisciplinary account in that it lied on the confluence of several literatures. The first is composed of managerial studies on the application of training to the workplace, including motivational processes as their central mediator. The second is composed of theoretical accounts on task and achievement motivation prominent in educational and social psychology. This confluence creates an interesting tension. It is obvious in reading the present study that ‘transfer’ is used in line with what Bransford and Schwartz (1999) have called the Direct Application theory of transfer. Certainly, there are other perspectives on transfer prevalent in other disciplines such as vocational education (Tuomi-Gröhn & Engeström, 2003), mathematics education (Lehtinen & Hannula, 2006), or computer-supported collaborative learning (Kapur & Kinzer, 2009). Many more could be named. The emphasis of the immediate use of what was learned in workplace applications is likely to be a result of the focus on the short-term benefits of training that can be found in many management and HRD contexts. In line with this view, motivation was seen as a single factor that can facilitate quick training application. We have criticized this view in the present study as being too coarse to account for the full complexity of intentional transfer processes It seemed to us that a way to overcome this problem was to use the best of both disciplines—HRD and learning sciences—to create a new account on motivation to transfer. Future research will have to show whether the proposed three-dimensional model is more successful in predicting transfer than a one-dimensional model.

This study revisited Noe's (1986) construct motivation to transfer. In a series of CFA analyses, it was shown that motivation to transfer has a three-factor structure consisting of autonomous motivation to transfer, controlled motivation to transfer, and intentions to transfer. This factor structure was built on a combined motivation theory approach unifying self-determination theory (educational psychology), expectancy theory (management), and the theory of planned behavior (social psychology). In a PLS-modeled path analysis, the study showed that social cues influenced controlled motivation, but not autonomous motivation; affective cues had an impact on both autonomous and controlled motivation. Intentions to transfer mediated the effect of autonomous motivation on training transfer. Compared to past transfer motivation research, this study tested several new directions that have theoretical implications. First, the motivational sequence explored indicates that intentional transfer processes cannot be captured with a single scale; rather motivational states and traits associated with social and affective cues can form an interdependent account of learner and environment (Hickey et al., 2001; Järvelä et al., 2009) This interdependency regulates motivation and has a differentiated effect on motivation and intention (Vauras et al., 2008). Second, the measurement time three months after training can be seen as a significant step toward developing an understanding of how motivation to transfer is affected when the learner interacts within her/his routine work environment, not affected by the training condition. More research in the temporal dimension of transfer motivation is needed, however, to overcome what was termed the dynamic problem of motivation to transfer (Gegenfurtner et al., 2009b). There are three limitations. First, the data consisted of self-report measures only. This was deemed appropriate for this study because the interest was in the trainees' subjective attitudinal, motivational, and intentional perceptions. However, we are aware of problems associated with leniency and self-serving bias that can likely occur in self-reported material. Second, study implications are constrained by characteristics of the population of trainees in occupational health education from which the sample was drawn. Third, generalization of the findings is limited to the time period three months after training. Since motivation to transfer is assumed to be a dynamic construct, all relationships explored must be considered with caution when generalizing to a different time frame.

## References


Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179-211.

Baard, P. P., Deci, E. L., & Ryan, R. M. (2004). Intrinsic need satisfaction: A motivational basis of performance and well-being in two work settings. *Journal of Applied Social Psychology*, *34*, 2045-2068.

Bentler, P. M. (2005). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of research on education*, *24*, 61-100.

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review*, *6*, 263-296.

Chin, W., & Newstead, P. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307-341). Thousand Oaks, CA: Sage.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt.

Festner, D. & Gruber, H. (2008). Conditions of work environments in fostering transfer of training. In S. Billett, C. Harteis, & A. Eteläpelto (Eds.), *Emerging perspectives of workplace learning* (pp. 215-231). Rotterdam: Sense.

Gegenfurtner, A., Festner, D., Gallenberger, W., Lehtinen, E., & Gruber, H. (2009). Predicting autonomous and controlled motivation to transfer training. *International Journal of Training and Development*, *13*, 124-138.

Gegenfurtner, A., & Gruber, H. (2009, August). *Do intentions mediate the transfer process? Testing the theory of planned behaviour*. Paper presented at the 13th Conference of the European Association for Research on Learning and Instruction, Amsterdam.

Gegenfurtner, A., Veermans, K., Festner, D., & Gruber, H. (2009). Motivation to transfer training: An integrative literature review. *Human Resource Development Review*, *8*, 403-423.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Hickey, D. T., Moore, A. L., & Pellegrino, J. W. (2001). The motivational and academic consequences of elementary mathematics environments: Do constructivist innovations and reforms make a difference? *American Educational Research Journal, 38*, 611-652

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Järvelä, S., Volet, S., & Järvenoja, H. (2010). Research on motivation in collaborative learning: Moving beyond the cognitive-situative divide and combining individual and social processes. *Educational Psychologist*, *45*, 15-27.

Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 75-170). Palo Alto, CA: Consulting Psychologists Press.

Kapur, M., & Kinzer, C. K. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported collaborative learning*, 4, 21-46.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

Lehtinen, E., & Hannula, M. M. (2006). Attentional processes, abstraction, and transfer in early mathematical development. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present and future trends. Fifteen essays in honour of Erik De Corte* (pp. 39-54). Oxford: Elsevier.

Leitl, J., & Zempel-Dohmen, J. (2006). Die Bedeutung des Arbeitsumfelds für die Veränderung der Transfermotivation. [The impact of work environment on the changing level of motivation to transfer] *Zeitschrift für Arbeits- und Organisationspsychologie*, *50*, 92-102.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-99.

Machin, M. A., & Fogarty, G. J. (1997). The effects of self-efficacy, motivation to transfer, and situational constraints on transfer intentions and transfer of training. *Performance Improvement Quarterly*, *10*, 98-115.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593-614.

Noe, R. A. (1986). Trainees' attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review*, *11*, 736-749.

Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology*, *39*, 497-523.

Pugh, K. J., & Bergin, D. A. (2006). Motivational influences on transfer. *Educational Psychologist*, 41, 147-160.

Ringle, C. M., Wende, S., & Will, A. (2005). *SmartPLS 2.0*. Available at http://www.smartpls.de.

Tai, W.-T. (2006). Effects of training framing, general self-efficacy and training motivation on trainees' training effectiveness. *Personnel Review*, *35*, 51-65.

Tuomi, Gröhn, T., & Engesröm, Y. (2003). Conceptualizing transfer: From standard notions to developmental perspectives. In T. Tuomi-Gröhn & Y. Engeström (Eds.), *Between school and work. New perspectives on transfer and boundary-crossing* (pp. 19-38). Oxford: Elsevier.

Tziner, A., Haccoun, R. R., & Kadish, A. (1991). Personal and situational characteristics of transfer of training improvement strategies. *Journal of Occupational Psychology*, *64*, 167-177.

Vauras, M., Salonen, P. & Kinnunen, R. (2008). Influences of group processes and interpersonal regulation on motivation, affect and achievement. In M. Maehr, S. Karabenick, & T. Urdan (Eds.) *Social psychological perspectives. Advances in motivation and achievement* (Vol. 15, pp. 275-314). New York: Emerald.

Velada, R., Caetano, A., Bates, R., & Holton, E. (2009). Learning transfer – validation of the learning transfer system inventory in Portugal. *Journal of European Industrial Training*, *33*, 635-656.

Vroom, V. H. (2005). On the origins of expectancy theory. In K. Smith & M. Hitt (Eds.), *Great minds in management: The process of theory development* (pp. 239-258). Oxford: Oxford University Press.

# Known Knowns and Unknown Knowns: Multiple Memory Routes to Improved Numerical Estimation

Dav Clark, Psychology Department, University of California, Berkeley, davclark@berkeley.edu
Michael Andrew Ranney, Graduate School of Education, University of California, Berkeley, ranney@berkeley.edu

**Abstract:** Conceptual change represents a crucial, challenging, learning component. This study hypothesized and observed evidence for two parallel forms of learning within the Numerically-Driven Inferencing (NDI) paradigm's rather minimalist intervention of providing direct feedback regarding a numerical estimate—feedback that yields remarkably robust cognitive alterations. The present experiment probed the nature of learning apropos recall or estimation improvements observed after participants (a) provided estimates, (b) received feedback, and (c) re-estimated after waiting for one day. The results show that improved estimation/recall was predicted by two independent elements—surprise at feedback and an explicit sense of episodic recall upon testing. This suggests at least two learning processes: (1) an explicit (though perhaps approximate) recollection of a quantity's magnitude and (2) a non-episodic semantic restructuring that correlates with surprise. Thus, even for concise, factual information, we educators might consider students' "unknown knowns"— knowledge that learners gain without any explicit understanding that they have done so.

## Introduction

Various streams of cognitive research suggest that significant conceptual change is difficult to effect (Chi, 2005; diSessa & Sherin, 1998; Lord, Ross, & Lepper, 1979). Numerically-Driven Inferencing (NDI) procedures, however, have yielded notable and encouragingly long-lasting levels of conceptual change with quite minimalist interventions (e.g., by providing estimators with a single, critical, highly germane, feedback statistic; cf. Rinne, Ranney & Lurie, 2006). The EPIC procedure represents one such intervention that is relatively compact and well specified (EPIC and NDI were introduced by Ranney, Cheng, Garcia de Osuna, & Nelson, 2001). Notably, EPIC has been shown to induce long-lasting conceptual change (e.g., Ranney et al., 2008), as evidenced by increased accuracy on estimations up to 12 weeks after the procedure (Munnich, Ranney & Bachman, 2005). If one can determine the cognitive factors that drive the efficacy of this particular curricular intervention, one might then be able to target these factors in developing and refining other pedagogical strategies.

In the EPIC procedure, participants engage with real-world numerical facts that bear on a societal issue, such as abortion, criminal justice, the environment, etc. (e.g., Garcia de Osuna, Ranney, & Nelson, 2004; Munnich, Ranney, Nelson, Garcia de Osuna, & Brazil, 2003). An example item/quantity is "the ratio of murders committed to prisoners executed in the U.S." People often poorly estimate these quantities, and thus the true values are surprising to many individuals. For the example above, individuals estimated anywhere from *two* to *a million* murders per execution (Munnich et al., 2003). (This broad range of innumeracy might be compared to order-of-magnitude errors made in reasoning about physics problems, although crime and physics likely diverge in terms of the complexity of, and one's familiarity with, their respective causalities.) Such starkly diverging estimators might have quite different beliefs regarding the effectiveness of capital punishment as a deterrent to murder (given that deterrence is commonly used in reasoning by both pro- and anti-execution participants).

Successful classroom curricula have been developed for improving students' estimation abilities as part of our laboratory's research on NDI—both in the contexts of high school mathematics and elite graduate journalism classrooms (Munnich, Ranney, & Appel, 2004; Ranney et al., 2008). The EPIC procedure is one component of these curricula in which participants (1) provide an **Estimate** for each policy-relevant item, (2) state what they would **Prefer** each quantity to be, (3) receive actual quantities as feedback to **Incorporate** (as new "**Information")**, and (4) indicate whether their preferences have **Changed** upon receiving feedback. Regarding change, our laboratory found that, after learning that about 250 murders occur per executed U.S. prisoner, participants significantly changed their preferences about the ratio (Munnich et al., 2003).

A fundamental question in cognition concerns the nature of what is learned. Some well-established psychological learning and memory models (Nadel & Moscovitch, 1997) would predict that changes in estimation accuracy must ultimately be mediated by episodic memory (which could give rise to abstracted semantic memories via a process of consolidation). In this case, we would expect participants' explicit memory for feedback (the "I" in EPIC) to correlate well with improvements in estimation accuracy at subsequent testing. Recent evidence suggests, however, that the re-modeling of existing conceptual structures may not depend on episodic memory formation (Tse et al., 2007). Indeed, there is broad evidence that learning may often be subserved by multiple memory systems, perhaps acting in parallel (Clark & Ivry, in press). In this case, we might expect increases in estimation accuracy even when participants report no memory whatsoever for the quantity provided as feedback—particularly if participants had pre-existing knowledge to support such learning.

Surprise upon receiving feedback provides evidence for pre-existing knowledge—specifically an incorrect prior expectation regarding the true value. Thus, surprise may correlate with the kind of non-episodic re-modeling described above. However, surprise may also reflect the emotional impact of the information (Munnich, Ranney & Song, 2007; Thagard, 2006). Therefore, it is important to assess not only surprise, but also whether the surprise had an emotional or intense character. In other words, surprise may mediate improved episodic memory, or it may indicate the existence of prior knowledge that may facilitate changes in semantic memory. Of course, these routes to improved estimation may operate partly or wholly in parallel.

While the hypotheses that follow are certainly not exhaustive, they constitute a set of potential reasons for the efficacy of the EPIC procedure. Most generally, engaging in estimation with feedback may result in a general increase in one's estimation ability (as Ranney et al., 2008, observed following a one-week, EPIC-related/inspired curriculum); such an effect would be seen even for items that received no feedback. We additionally considered that learning would be driven by surprise (e.g., Munnich et al., 2005). Moreover, improvements in estimation could be driven by a direct (potentially approximate) episodic memory of feedback. If *all* improvement were driven by episodic memory, though, we would expect little extra power in a model that included both surprise and episodic memory ratings beyond the power of a model that only included one of these. If, however, a model including both surprise and memory provides a significantly better fit, this offers some indirect evidence that at least a portion of improvement is occurring through a non-episodic route.

We additionally considered the optimal timing for feedback: Immediate engagement with feedback following the generation of an incorrect estimation may (a) interfere with the successful learning of the new, correct, immediately provided information. However, another possibility regarding immediate feedback is that it may (b) help encourage an individual to both engage with and correct her beliefs. Thus, if we were to find that participants improve more with *delayed* feedback than immediate feedback, hypothesis (a) would be supported. Alternatively, if performance were superior with immediate feedback, (b) would seem more plausible.

## Experimental Methods

The following experiment was designed to assess whether estimative improvement occurs even with respect to items for which no feedback was received—as was found in curricular NDI studies (e.g., Munnich et al., 2004; Ranney et al., 2008). The experiment addresses (1) the effects of the timing of feedback on subsequent improvements in numerical estimation—as well as (2) whether these improvements are necessarily mediated by episodic memory. A subset of the EPIC procedure was used to explore these issues; participants engaged only in estimation ("E") and feedback ("I"), leaving aside personal preference (“P” and “C”).

### Participants

Twelve people (seven female) participated (and 19 participated in a later experiment that replicated our main findings), including University of California, Berkeley, undergraduates and members of the general public recruited via online recruitment systems (RPP and RSVP). They received either course credit or $20 for their participation in two one-hour sessions over two consecutive days. Ages ranged from 18-56 years.

### Materials

Numerical facts (106 of them) were selected from Ranney et al.'s (2008) collection (See the introduction for an example). Three statistical facts were set aside for the basis of example items (namely US population, world population, and US Gross National Income). Items ranged over a number of topics, and included politics, population dynamics, economics, the environment, education, crime etc. Most items were expressed in percentage form, with the rest being counts of dollars, people, events, or things. For numbers above 999, a comma was used, as in "13,600." For numbers in the millions, billions, or trillions, the appropriate word was used to indicate the order of magnitude (e.g., "300 million"). This was intended to minimize possible confusions about the exact value of the number.

### Procedure

Custom software utilizing Vision Egg (Straw, 2008) presented all materials and collected responses (source code available upon request). Descriptions of numerical facts were presented in four or fewer lines of text (with fewer than 55 characters per line). A prompt for numeric entry was located below the description. Feedback concerning the veridical value was provided in a third location, between the description and the text-entry area.

#### Blocks of Items

Items were randomly distributed into the following four kinds of blocks. Each of these blocks was involved in two or more phases over the course of the experiment. **E**: Participants *only* provided **Estimates** in a single phase. **EI**: Participants provided **Estimates,** followed *immediately* by correct numerical **Information** as feedback (i.e., feedback was provided in the same phase as the initial estimation). **E_I**: Participants provided **Estimates**, and then received correct numerical **Information** in a phase that was *well-separated* (but by less than 45 minutes) from the phase in which they provided their Estimate (i.e., "_" signifies a temporal delay). **New**: A block of items was reserved to provide a gauge of false recognition or false recollection.

## Experimental Phases

Participants engaged in a number of self-paced phases on each of the two consecutive days, as Figure 1 depicts. The structure of stimulus presentation and response collection was uniform across a given phase. During the first day, analogous to a "study" period, participants completed three partially similar phases of numerical estimation and/or informative feedback. (See Phases 1-3 in Figure 1.) The second day was analogous to a "test" period, in which participants' learning was assessed (Phases 4-7 in Figure 1).

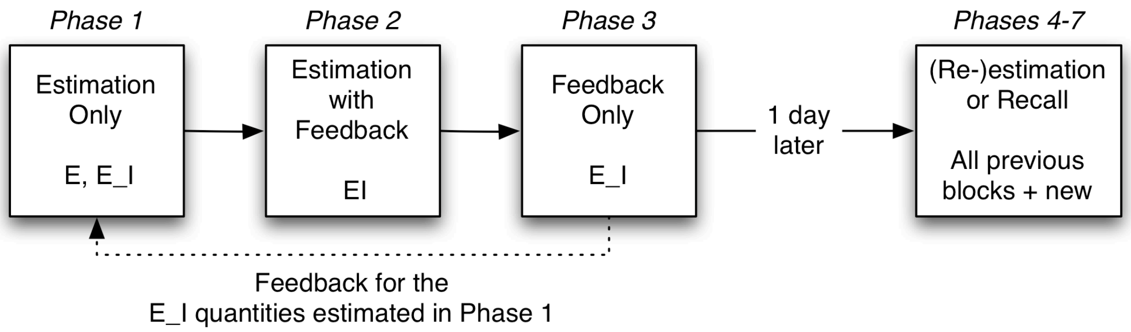



Figure 1. A schematic of the experiment's seven phases. *Phase 1*: Estimates were obtained for the E and E_I blocks of items (23 each), randomly intermixed in one phase. *Phase 2*: Participants provided 23 Estimates that were immediately followed by Informing the participant of the correct value. *Phase 3*: Feedback (I) was provided for the 23 items from the E_I block that had been Estimated in Phase 1. *Phases 4-7*: Subjects estimated (or recalled) quantities and provided explicit memory ratings for all previous items as well as 34 new items.

During estimation (Phases 1 and 2, with 23 items each), subjects were given a textual description of an item's quantity, followed by a prompt to provide an estimate. For Phase 2, feedback was provided 500 milliseconds after each estimate was entered. For Phase 3 (with 23 items), the correct numerical value was provided *prior* to the textual description in order to minimize covert estimation.

In Phases 2 and 3 (thus, for blocks including "I"), participants provided a self-report on their subjective sense of surprise. Their responses were restricted to the following three choices: (1) Little or no surprise, (2) Genuine surprise, or (3)"Visceral" or intense surprise. The presence of any form of surprise is an indication of pre-existing memory structures, while the distinction between levels (2) and (3) may indicate a difference in the emotional impact of the feedback for that item. (Note that Munnich, Ranney, & Song, 2007, found that a prospective measure of surprise reliably predicted the sort of retrospective surprise ratings solicited here.)

On Day 2, trials were similar to the estimation-only trials in Phase 1 described above in that no additional feedback was provided. An additional 34 items from the "new" block were randomly intermixed with the items presented during study. Furthermore, participants rated their memory for the item according to the following four levels: (1) "The item is new to me," (2) "The item was presented yesterday, but I have no sense of the value provided as feedback," (3) "The item was presented yesterday, and I have some sense of the correct value," or (4) "The item was presented yesterday, and I have a fairly accurate recollection of the value."

Choice 1 indicates no recognition or recollection. This is equivalent to labeling the item as "new," and it is the correct response for items from the new block. As a group, choices 2-4 indicate that the item is "old," but with varying levels of familiarity and/or recall. These are correct responses for the E, EI and E_I blocks (although choices 3 and 4 entail a belief that the participant actually received feedback at study, and so might also be considered incorrect for the E block). Choices 2 and 3 indicate perceived recognition, but at least a partial failure in explicit recall. Choice 4 indicates a subjective sense of fairly complete recall.

Explicit recall is used herein in a somewhat different way than in most learning and memory studies. Indeed, these memory ratings can be viewed as a form of metacognition regarding the estimation process.

## Analysis

We modeled improvement as a binomial outcome (as did Munnich et al., 2005). This allows for the treatment of items that have differing distributions within a unified framework. (It would be difficult modeling both percentages and values in, say, the billions, particularly given our sample size.) Items were labeled as to whether estimates improved or not. These labels were fit with a binomial generalized linear model, using the lme4 package in R. This treatment allows for a full multi-factorial mixed-effects analysis. Below, participants are always included as a random effect, and other factors are treated as fixed effects. Contrasts were evaluated using the multcomp package, which controls for family-wise error rate (Hothorn, Bretz, & Westfall, 2008).

Unless otherwise noted, data were pre-processed to remove ties. This was done to allow for a null hypothesis that 50% of the remaining items randomly improved and 50% randomly worsened. If we counted ties as failures to improve, then random drift would end up spuriously suggesting the lack of an effect. Removing ties allowed for tests of whether estimates, on average, improved more than they worsened—both formally and when examining graphs. Otherwise, the removal had little effect on the results.

## Results

### Improvements in Participants' Estimation Accuracy and/or Numerical Recall

Figure 2 shows the proportions for the number of items showing improvement (to any degree), by condition. We can easily reject a null model (i.e., with all conditions modeled by the same mean) in favor of a model including the three feedback conditions ($\chi^2(2) = 25.9$, $p < 10^{-6}$). Post-hoc comparisons between each condition and chance levels, as well as between condition comparisons (as in a Tukey HSD test) were performed simultaneously. In the no-feedback condition (E), estimation improvement did not differ significantly from chance ($p = .39$), although improvement with Immediate (EI) and Delayed (E_I) feedback were clearly above chance ($p < 10^{-4}$). This may seem quite expected, but it *might* have been the case that improvements were at least partially driven by general improvements in estimation skill, and this would have led to at least some modest improvements even without feedback on test items. Indeed, this kind of estimation skill development was the successfully accomplished goal of various EPIC-based curricula (e.g., Munnich et al., 2004; Ranney et al., 2008). In the present, less extensive, experimental manipulation, though, we understandably elicit no such skill improvements. Thus, we assume that these improvements are driven almost entirely by item-specific learning.


Figure 2. Proportion of items from each block for which estimates improved on Day 2 (as compared to estimates on Day 1). Items with identical estimates on Day 1 and Day 2 have been removed; thus chance is .5. Asterisks note significant differences.

Contrary to part of our hypotheses, participants exhibited essentially equal proportions of estimation improvements for the EI and E_I blocks. While these treatments yielded significantly more improvement than for the no-feedback case (both $p$'s $< 10^{-4}$), they did not differ significantly from each other ($p = .79$).

### Recollection and Surprise

As is often the case, the participants' forced familiarity judgments appeared to be superior to their own assessments of their memory (Merikle, 2007). In participant debriefings, several individuals claimed to be uncertain as to whether items were old even from Phase 1 to Phase 3 for items in the E_I block—over an interval of less than 45 minutes! However, participants performed well at discriminating between old and new items a day later when given a forced choice; 76% of new items were identified as new on Day 2, compared to an average of less than 9% regarding previously seen items. This level of recognition accuracy is not surprising, given the considerable depth of processing involved—and the rich, pre-existing, memory structures available for scaffolding these episodes. (This may be another point of departure regarding items such as the likelihood of a murder ending in an execution, in contrast to the pre-existing memory structures K-16 students might hold regarding the physics of mechanics or electromagnetism. In both cases, though, it seems probable that multiple conceptual changes may result—e.g., "Wow, the high murders-per-execution ratio makes me think [a] that gangsters would rarely fear the death penalty, given [b] how rarely their fellow gang members are executed and [c] that death rows are basically blossoming geriatric wards, meaning [d] that gangsters would be more likely to bring guns." Such conceptual changes do not seem wildly different, in many important respects, than those required to understand that ballistic trajectories are curvilinear, rather than diagonally rectilinear.)

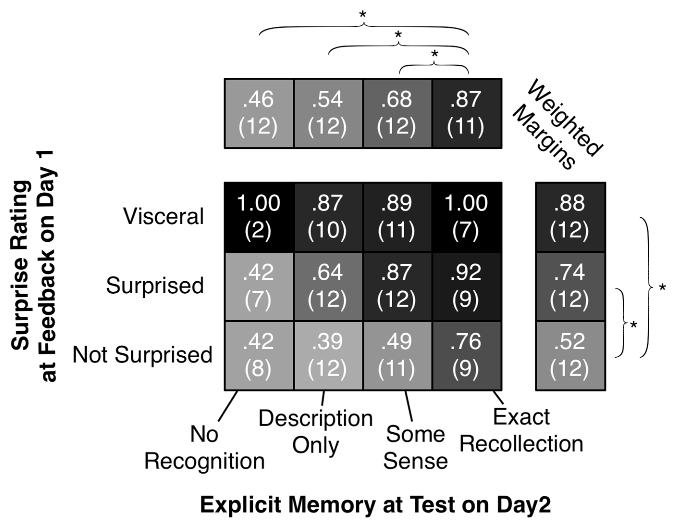



Figure 3. Proportions of items with improved estimation accuracy (or recall) on Day 2. The effects of surprise (Day 1) and explicit memory rating (Day 2) are shown independently via weighted margins, and (the lack of) an interaction via the central table. Asterisks note individual factors' significant differences; significant interactions were not observed. Only items with specific feedback are included (i.e., EI and E_I conditions). Proportions represent between-subject means, and the number of subjects contributing to each cell is shown in parentheses.

As the no-feedback condition (E) yielded non-significant changes in estimation accuracy, we hereafter consider only items from conditions including feedback ("I"; i.e., EI and E_I). These effects are depicted in Figure 3. We find that a model with both surprise and declarative memory responses cannot be rejected in favor of a reduced model excluding memory ($\chi^2(3) = 34.8$, $p < 10^{-7}$), and the comparison involving the removal of surprise is even more striking ($\chi^2(2) = 295.22$, $p < 10^{-16}$). An inclusion of an interaction term does not yield a significantly superior model ($\chi^2(6) = 2.85$, $p = .8$), nor does the inclusion of the feedback condition (EI vs. E_I; $\chi^2(1) = .8$, $p = .36$). There was a small but non-significant difference between surprise ratings for EI and E_I blocks: subjects rated 65% of EI block items as surprising ("2" or "3") vs. 59% for E_I. No straightforward trend was observed with explicit recollection. The timing of feedback may have some effect on estimation improvement that is mediated by surprise, but such issues seem best addressed in subsequent research. Below, we only consider comparisons within surprise level and memory ratings independent from one another.

"Exact recall" for the value of a given item (i.e., memory response "4")—as compared to other memorial impressions—was a highly significant predictor of improved estimation ($p$'s < .001 for ratings "1" and "2;" $p = .01$ when compared with response "3"). No other comparisons between memory levels were significant. For surprise, items with moderate and visceral ratings were significantly more likely to exhibit improvement than those with a no-surprise rating ($p < .002$ in both cases), but did not differ significantly from one another. According to our earlier logic, this suggests more of an effect of prior knowledge than emotional impact, although emotional impacts may still play nontrivial roles. Note that participants provided the exact numerical figure given as feedback only 35% of the time when selecting choice 4. Even if we broaden this accuracy criterion liberally to items for which participants are within 15% of the true value, they were only correct about 74% of the time.

Finally, if we consider the link between surprise and explicit recall, there seemed to have been little relationship in the present study (as opposed to some other studies). The correlation of fixed effects between memory and surprise terms in our model was consistently smaller in magnitude than .1. This, combined with the lack of a significant interaction term, provides some evidence for independent learning processes.

## Exclusions

As many as three items lacked estimates from some subjects or exhibited a clear lack of understanding in a response (e.g., a "10 million" response for a question soliciting a percentage) and these items were excluded from the analyses above. Due to a technical issue, one participant did not receive the standard E manipulation, but was included in memory and surprise-related analyses, as these analyses did not include E trials.

## Discussion

Given the overall improvements in estimation ability evidenced in curricular studies by Munnich et al. (2004) and Ranney et al. (2008), it is of interest that we see no statistically significant improvement in items that didn't receive feedback (the "E" block). Nonetheless, it seems that learning in this considerably shorter present experiment was largely item-specific and related to the integration of feedback. This lack of improvement in the present experiment may be due to a lack of time for reflection or the development of strategies—which were highlighted, taught, and fostered in the curricular studies. (Munnich et al., 2004, and Ranney et al., 2008, also focused, to a fair degree, on preferences and personalized policies.)

In addition, we had originally hypothesized that the EI block would yield the greatest overall improvement in estimation. We assumed that immediate feedback would be more memorable because participants would have the opportunity to directly compare their estimate with the actual value and consider its validity. This would have arguably led to a more tangible interaction with the true value than in the E_I condition. However, our data show that the E_I block, if anything yielded larger improvements. (Although EI and E_I conditions did not significantly differ, this pattern has held up in subsequent experiments as well.) This may be due to a number of reasons, warranting further experimental inquiry.

One possibility is that we are observing something akin to a distributed learning effect—because even though the feedback is provided only once, the size of the effect is reminiscent of the effects of well-spaced study periods reported by Pashler et al. (2007). Note that this is a major departure from standard spacing effects, as the correct information is presented *only once.* However, while EI items were only shown once before testing, item descriptions in the E_I condition were shown twice—initially in Phase 1 (when participants provided their initial estimates) and again in Phase 3 (when participants were given feedback). This may have enhanced those items' representations (e.g., with a bit more time-on-task), providing a richer or more stable context for incorporating feedback (as with Tse et al., 2007). On the other hand, these data might also suggest that we have not explored the most relevant timescales. For example, the optimal estimate-feedback delay may be much shorter, as with the work of Johnson and Siefert (1994). A final possibility is that the participants' assessments of surprise at receiving feedback may inappropriately draw their attention back to their incorrect prior knowledge. Thus, we might find generally enhanced learning and larger effects of feedback spacing, were we to encourage participants to engage more fully with the new information—perhaps as is done in the usual EPIC procedure, in which participants revise their preferences/policies based on the new information and the inferencing that such feedback triggers (as in Munnich et al., 2004, and Ranney et al., 2008).

## Learning Without Recall

From the point of view of a memory theory, the most interesting result is perhaps the existence of learning even when participants claimed "no sense" of the numerical value provided at feedback—rather like a memorial analog to blindsight (Merikle, 2007). This argues against the notion that improvements in estimation are simply the result of explicit episodic memory. The result is reminiscent of extant *dual-process* memory models. For example, Davachi et al. (2003) suggest that successful recognition could occur through a process of recollection and/or a sense of familiarity. These processes moreover appear to be subserved by distinct sub-regions in the medial temporal lobe. In the present study, we see improvement in numerical estimation—which is perhaps most akin to a cued recall task for EI and E_I items—without full recall of the number presented on the previous day. Thus, the task here is perhaps more naturally expressed in the language of the *remember/know* distinction (Knowlton, 1998). That is, while participants appear not to *remember* a number from the previous day, there is still a sense in which they *know* the number better than they knew it the day before.

Based on the significance of the existing results, however, it seems reasonable to posit that a non-episodic form of learning undergirds some of the improvement in participants' abilities to estimate accurately. Further, the learning for improved estimation (or memory) seems to occur even without an explicit recollection of the feedback from the prior day. This argues for some implicit and/or rapidly semanticized learning in support of these improvements.

## Educational Implications

We have provided support for the effectiveness of instruction that engages with a person's pre-existing knowledge. A finding of central importance to educators is the lack of improvement on E items. It seems that if the pedagogical goal is to improve numeracy, then simple engagement with a given set of quantities is not enough to enhance estimation abilities for other quantities. Item specific training could be augmented with more generalizable strategies, though, as has been demonstrated with other interventions (e.g., Munnich et al., 2004). Students' abilities to estimate quantities for which they *have* received feedback can be quite good, and it seems that there is at best a small effect of whether students think about their own beliefs immediately preceding or a day prior to receiving correct instruction. This effect is in line with results from elements of the distributed learning literature (Pashler et al., 2007)—perhaps with some form of priming for subsequent encoding occurring during one's initial estimation. Future studies will examine differing intervals for both feedback and retention. Given our results, though, it is clear that effective learning can occur over a panoply of semantic, statistical, items in a framework that includes engaging students' pre-existing understandings of the material. It is worth noting further that, unlike some other forms of conceptual change cited above, the nature of what was learned here may be of a more continuous nature than, say, "what determines the phases of the moon."

Such information may be more amenable to "implicit" forms of learning than concepts that have a more categorical, intricate, or non-linear structure. It bears considering, though, whether one can introduce more graded forms of learning into the logical structures of science and related fields.

In general, participants displayed fairly good recognition of items they had seen previously, distinguishing new items from ones presented the previous day. However, improvements occurred frequently even when participants claimed "no sense" of the feedback, particularly when they found the item surprising—in other words, when estimation improvement was predicted more by surprise than by explicit recall. In such cases, learners seem to have poor metacognitive inclinations—they do not "know what they know."

Combined with strategies for evaluating one's own estimates (e.g., those provided in curricula by Munnich et al., 2004, and Ranney et al., 2008), a learner could be encouraged to generate what they believe to be a "guess," but which might turn out to be well informed by what they have in fact learned. If encouraged to make such guesses, students should still spend time reasoning about their guess; prior work from our group (Ganpule, 2005) has indicated that good estimators spend more time representing a quantity before estimating it, consistent with work from the problem solving literature. Guessing too quickly is more often counterproductive. Indeed, this result encouraged the curricula of Munnich et al. (2004) and Ranney et al. (2008) to highlight the importance of self-critique and disconfirmation. (However, one of us anecdotally finds people, at times, to be markedly underconfident estimators. For instance, people frequently claim to have "no idea" of the population of California, yet often laugh out loud when asked if it could be "1,000.") Another idea is that students might infer something about the quality of their estimate by introspecting on whether they remember being surprised at a number during instruction. Such surprise could indicate an improvement of one's sense of magnitude.

More generally, participants were able to learn something in this experiment without being aware of their learning. This strongly suggests that non-episodic (i.e., implicit or rapidly semanticized) learning processes could be involved even in the development of "declarative" factual information. Thus, depending upon an instructor's particular pedagogical goals, one might proceed with instruction without concern for students' initial awareness of their improved sense of certain facts (perhaps as in "immersion" language-learning experiences). Over the course of an intervention, one would hope that students would eventually come to know and trust their new knowledge on the topic, but such metacognitive awareness needn't come lock-step with improvements in the more basic knowledge. This line of reasoning is reminiscent of studies of children applying abstract mathematical rules before they are aware of doing so (Siegler, 2000).

Of course there is also a clear role for explicit episodic recall in learning numerical information. In particular, when participants displayed zero or moderate amounts of surprise, improvements in estimation/recollection were likely only if they believed they could recall the number. Thus, when students lack sufficient schemas for incorporating the numerical knowledge and the information imparted is unsurprising, rote memorization may be virtually the only remaining route. Of course, a complete pedagogy could include the construction of knowledge structures that would then allow for the parallel recruitment of non-episodic learning.

Instructional materials that elicit surprise in students may allow such students to learn without conscious awareness that they have learned anything—at least in domains that are scaffolded by nontrivial preexisting knowledge. If the material is unsurprising, it appears that episodic encoding may be a critical step in successful improvement. It should be noted that "surprise" might be often used as too specific a notion. It may be that the relevant feature has more to do with general emotional, motivational, or inspirational, salience—or how interesting the material is to students. (See Kang et al., 2009, on connections between surprise, motivation, and curiosity, the latter two of which are likely enhanced by soliciting estimates—and even preferences—as in the EPIC procedure). Certainly, however, it seems that there are multiple routes to learning even relatively concise facts, and a successful pedagogy might usefully engage factors such as surprise and engagement with pre-existing knowledge to bolster more rote forms of learning.

## References

Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences*, *14*(2), 161–199.

Clark, D., & Ivry, R. B. (in press). Multiple systems for motor skill learning. *Wiley Interdisciplinary Reviews: Cognitive Science*.

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences*, *100*(4), 2157–2162. doi:10.1073/pnas.0337195100

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, *20*(10), 1155–1191.

Ganpule, S. (2005). *Strategy use in numerical estimation: Investigating the effects of an EPIC curriculum.* Unpublished master's thesis, University of California, Berkeley.

Garcia de Osuna, J., Ranney, M., & Nelson, J. (2004). Qualitative and quantitative effects of surprise:(Mis) estimates, rationales, and feedback-induced preference changes while considering abortion. In K.

Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 422-427). Mahwah, NJ: Erlbaum.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology Learning Memory and Cognition*, *20*, 1420-1420.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), 963-973. doi:10.1111/j.1467-9280.2009.02402.x

Knowlton, B. J. (1998). The relationship between remembering and knowing: A cognitive neuroscience perspective. *Acta psychologica*, *98*(2-3), 253-265.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109.

Merikle, P. (2007). Preconscious processing. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness*. Malden, MA: Wiley-Blackwell.

Munnich, E., Ranney, M., Nelson, J., Garcia de Osuna, J., & Brazil, N. (2003). Policy shift through Numerically-Driven Inferencing: An EPIC experiment about when base rates matter. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society* (pp. 834–839). Mahwah, NJ: Erlbaum.

Munnich, E., Ranney, M. A., & Song, M. (2007). Surprise, surprise: The role of surprising numerical feedback in belief change. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the twenty-ninth annual conference of the cognitive science society* (pp. 503–508). Mahwah, NJ: Erlbaum.

Munnich, E. L., Ranney, M. A., & Appel, D. M. (2004). Numerically-Driven Inferencing in instruction: The relatively broad transfer of estimation skills. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society* (pp. 987–992). Mahwah, NJ: Erlbaum.

Munnich, E. L., Ranney, M. A., & Bachman, M. L. N. (2005). The longevities of policy-shifts and memories due to single feedback numbers. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-seventh Annual Conference of the Cognitive Science Society* (pp. 1553–1558). Mahwah, NJ: Erlbaum.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin and Review*, *14*(2), 187–193.

Ranney, M., Cheng, F., Garcia de Osuna, J., & Nelson, J. (2001). *Numerically driven inferencing: A new paradigm for examining judgments, decisions, and policies involving base rates*. Paper presented at the Annual Meeting of the Society for Judgment & Decision Making.

Ranney, M. A., Rinne, L., Yarnall, L., Munnich, E., Miratrix, L., & Schank, P. (2008). Designing and assessing numeracy training for journalists: Toward improving quantitative reasoning among media consumers. In P. Kirschner, F. Prins, V. Jonker, & G. Kanselaar (Eds.), *International Perspectives in the Learning Sciences: Proceedings of the Eighth International Conference for the Learning Sciences, Vol. 2* (pp. 2-246 to 2-253). International Society of the Learning Sciences, Inc.

Rinne, L. F., Ranney, M. A., & Lurie, N. H. (2006). Estimation as a catalyst for numeracy: Micro-interventions that increase the use of numerical information in decision-making. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings of the seventh international conference on Learning sciences* (pp. 571-577). Mahwah, NJ: Erlbaum.

Siegler, R. S. (2000). Unconscious insights. *Current Directions in Psychological Science*, *9*, 79-83.

Straw, A. D. (2008). Vision egg: an open-source library for realtime visual stimulus generation. *Frontiers in neuroinformatics*, *2*. doi:10.3389/neuro.11.004.2008

Thagard, P. (2006). *Hot Thought: Mechanisms and Applications of Emotional Cognition*. The MIT Press.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., et al. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76-82. doi:10.1126/science.1135935

## Acknowledgments

We thank Tawny Tsang above all others for myriad forms of assistance, Ed Munnich for extensive comments, as well as Andrew Galpern and other members of the Reasoning Group at UC Berkeley. Thanks also to Tom Wickens for statistical advice, Rich Ivry for his overall wisdom and support, and Mick Rugg and Lila Davachi for their input on experimental design issues. Finally, we thank UC-Berkeley's Committee on Research.

# The Role of Concretization in Acquiring Design Knowledge

Tamar Ronen Fuhrmann, tamarrf@gmail.com
Yael Kali, yaelk@technion.ac.il

**Abstract:** This research explored the nature of design knowledge by examining the processes in which graduate students in education learned to design educational technologies. We developed two rubrics to assess: (a) the degree to which students were able to translate their design ideas into concrete design artifacts (concretization rubric), and (b) the degree to which they designed artifacts that followed a socio-constructivist pedagogical approach versus a teacher-centered transmissionst model (epistemology rubric). Outcomes indicated that as students developed their concretization skills, they were able to become aware to and reduce gaps between their "theoretical" and "applied" epistemologies. By making their design ideas more concrete, students were able to carry out productive negotiations about these ideas with instructors and peers, and to explore them in relation to theory and to expert design knowledge.

## Rationale

Research in the learning sciences has shown that many opportunities to learn arise in the course of designing an artifact in general, and in designing an artifact intended for others to learn with, in particular (Papert, 1991). The potential of designing as a process that supports learning has been documented for a wide range of ages and levels of expertise. For instance, Harel (1991) explored the learning that takes place when fourth grade children develop mathematical software products for other students in their school. She showed that the young designers learned not only about mathematics (fractions) and programming (Logo), but also about design and user interface. Kafai (2006) showed similar outcomes with fifth grade children who designed and developed computer games for their peers. She argues that: "The greatest learning benefit remains reserved for those engaged in the design process, the game designers, and not those at the receiving end, the game players" (p. 39). The impact of engaging students in design processes on their learning was also found with middle school students; for instance, Kolodner et al. (2003) indicated that the Learning By Design approach significantly enhanced middle school students' motivation, their collaboration and metacognitive skills, and their scientific understanding in topics included in the products they developed (earth and life sciences).

In this research we explore the learning that occurs in a design process with a target audience that received only little attention in the learning by design literature, namely, graduate students in education. Researchers in various disciplines, such as the sciences of *learning*, *instruction* and *design* have recently begun to synthesize practical and theoretical knowledge regarding how to guide and inspire creators of innovative educational tools (DiGiano, Goldman, & Chorost, 2009). However, more research is required to better understand the ways in which graduate students in education—who are potential educators, curriculum-designers, learning-scientists, or policy makers—develop skills in designing technology-based curriculum modules, and how they can be supported in this process.

In an earlier study, Ronen-Fuhrmann, Kali, and Hoadley (2008), showed that there is an important added value in engaging graduate students in designing their own technology-based curriculum modules; while working on their design projects, students became more aware of gaps between what was defined as their "theoretical epistemologies about learning" (ideas expressed during general discussions about design, usually representing a socio-constructivist approach) and their "applied epistemologies about learning" (ideas reflected in artifacts they created, which tended to apply more transmissionist approaches), and were able to reduce these gaps. In this manner, students' epistemologies about learning became more coherent – an important outcome for students in education, whether or not they intend to design curriculum materials. In the current study we focus on the nature of knowledge and skills that students gain in educational technology design courses – a type of knowledge which the literature generally refers to as design knowledge. We focus on a specific aspect of design knowledge found in this research – the ability to concretize abstract design ideas, and explore the relationship between students' development of concretization skills and their ability to reduce their epistemological gaps.

## Research Settings

This research investigated the learning that took place in a semester-long course named "*Designing Educational Technologies*" designated for graduate students in science education, which was designed, developed and instructed by the authors of this paper. Students in the course were introduced with theoretical and practical aspects of educational technology design; they worked in groups of two or three to design a mockup of an educational technology module, but were not required to develop or implement these modules. The rationale was to focus on exploring students' learning while designing, without the constraints caused by any development tool, and without requiring a time-consuming development process.

The course was based on an instructional model, which evolved in an iterative design process in a previous study (Ronen-Fuhrmann et al., 2008). The final version of the model includes three main elements, which reflect a unique application and integration of three frameworks: (a) the ADDIE structure (Analyze, Design Develop, Implement, Evaluate) (Dick, Carey, & Carey, 2001). (b) the studio approach to instruction (Hoadley & Kim, 2003; Schon, 1983), and (c) the use of the Design Principles Database (DPD) (http://edu-design-principles.org), a web-based resource of socio-constructivist pedagogical design-principles, which were written by design experts (Kali, 2006; Kali & Linn, 2007; Kali, Linn, & Roseman, 2008).

## Methodology

We used a case-study methodology to examine students' learning processes and their development of design knowledge throughout the course. A collective case-study approach–often referred to as "multiple case study" (Stake, 1994)—was implemented. This approach is aimed at providing insights into an issue or problem or to refine a theory by exploring similarities and patterns between several case-studies. In this research, each group of 2-3 students, who worked on one design project during the course, was defined as a case-study.

The study was conducted with 14 groups (33 graduate students) who participated in three enactments of the *Designing Educational Technologies* course between the years 2005-2007. Most students had some experience in teaching or were active science teachers. They had some experience in designing curricula but most of them had no experience in designing technology-based learning modules. In order to characterize student learning in each of the iterations the following data sources were used (Table 1).

Table 1: Data sources.

| Data source | Description |
|---|---|
| Likert type surveys | At the end of each enactment students were asked to evaluate various elements of the course (such as the structuring of the design process, working with peers, using the DPD etc.) on a 1 to 5 scale. |
| Reflective essay | At the end of each enactment students were required to write a reflective assay about their design process. |
| Semi-structured interviews | At the end of each enactment we conducted interviews with two students who were asked to reflect about their design process. |
| Records of online discussions | Whole class online discussions about the literature and group online discussions regarding the design of the group's module were automatically recorded at the courses' website. |
| Student artifacts | During the semester we collected documents produced at various stages of the course. These documents included formal design artifacts students were required to write, as well as informal notes and sketches students created to discuss their ideas with peers and with us. |
| Reflective journal | Following each class-meeting we documented events related to each of the groups' progress, the discussions we had with students, or other events that seemed relevant for analyzing each group's learning processes. |

These data were analyzed using two rubrics; the first, entitled "concretization rubric" was used to evaluate the degree to which students were able to translate their design ideas into design artifacts; the second, entitled "epistemology rubric" was used to examine the epistemological changes that students went through during the course. We explain our rationale for developing these rubrics, and the way we used them in the sections below.

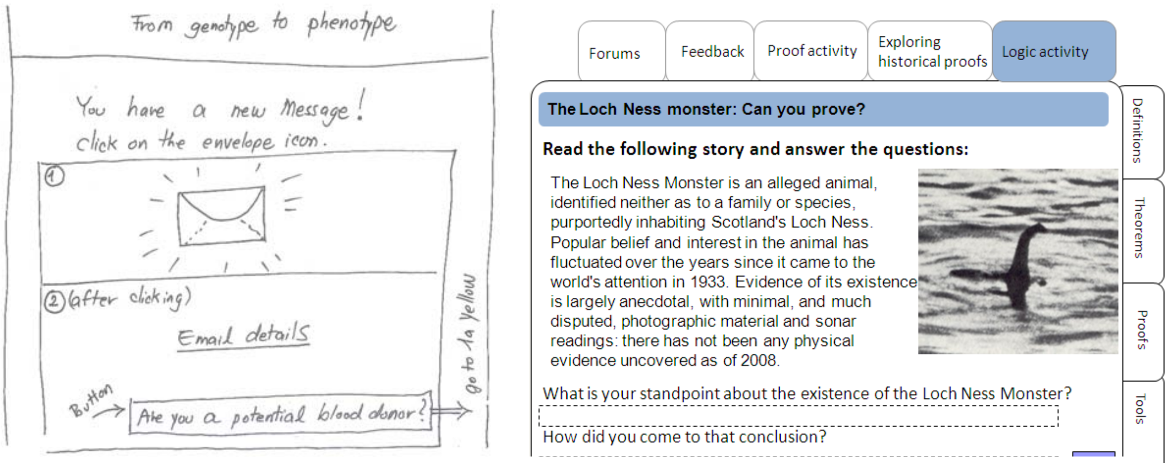
### Concretization Rubric

From early stages of the research we noticed that some students were able to easily translate their design ideas into concrete artifacts, while other students found this process very difficult. Following our preliminary observations, we decided to focus our assessment of design artifacts created by students using this lens. We developed a rubric (Table 2) which describes six levels of concretization of design ideas. Each level is associated with a stage in the design process in which a higher level of concretization is required. Thus, level 0 represents a stage in which students are required to discuss their module only in a theoretical manner. Level 5 represents the highest level of concretization; an artifact at this level should depict a full learning environment with a clear navigation system illustrating sequences of activities with clear instructions for learners.

It is important to note that although each stage in the instructional model was designed to provide students with the design skills required at that stage, including guidelines for concretization, we did not anticipate a one-to-one relationship between the knowledge taught and the knowledge gained by students at each stage. We also want to stress that we do not view concretization as a goal of the design process. Rather, we refer to concretization as important means to reach pedagogically sound artifacts.

Table 2: Rubric for assessing concretization of design ideas.

| *Level of concretization* | *Design Artifact Characteristics* | *Example Student Expressions and Artifacts* | *Required in Stage* |
|---|---|---|---|
| 0 – Theoretical knowledge | No design artifact. Only theoretical sayings regarding a planned module. | "It's very important to build activities that would be relevant and interesting to the learner" | Analysis |
| 1 – Collection of design ideas | A collection of design ideas for the module. The ideas only generally refer to the way a learner might act in the module. | "Learning throughout the whole module should follow a specific inquiry question". (Excerpt from a discussion of one of the groups regarding their design of a biology module. They planned to design the activities around an inquiry question but were not concerned at this stage about the nature of this question). | Brainstorm Activities |
| 2 – Initial activity sequence | Graphical or verbal description of a set of activities, with an indication of which activity should take place before or after another. | "First we should show them [the learners] the story about the family tree, then have them review the algorithm for scanning the tree, and then use the simulation" (Excerpt from a discussion regarding the design of a technology for high-school computer-science learners) | Build Flow |
| 3 – Initial translation into features | Ideas are translated to actual features and presented in a way that shows how a learner might act in the module. | Figure 1a shows a sketch of the way students envisioned an activity they planned for a module in genetics. Learners in this activity were required to decide whether they can donate blood to a kid with cystic fibrosis. At this stage this was the only activity they developed in their module. | Design Features |
| 4 – Initial learning environment | Mockup of the module showing some of the activities, with instructions for learners. An initial navigation system is present. | (See Figure 1b showing design knowledge level 5 – full learning environment as reference). | Mockup: Iteration 1 |
| 5 – Full learning environment | Mockup of the module showing most of the activities with clear instructions for learners. A clear navigating system is represented. | Figure 1b shows one screen from a mockup of a module designed for teaching logical thinking for middle school math students. The buttons at the top and side of the screen indicate that this screen is part of the whole learning environment. Each of these buttons was linked to a screen in the mockup, which was developed in a similar level of detail. | Mockup: Iteration 2 |




(a) Design artifact showing level 3 of concretization (b) Design artifact showing level 5 of concretization

Figure 1. Examples of artifacts showing levels 3 and 5 of concretization.

## Epistemology Rubric

The term "epistemology" in the learning sciences has traditionally referred to several entities in the nature of knowing, such as beliefs, conceptual understanding, perceptions, feelings and emotions. Since our analysis did not enable us to fully distinguish between these entities, we decided to use the general term "epistemology" in order to refer to any idea regarding how people learn expressed by the students, in their sayings and doings.

As mentioned above, a gap between students' "theoretical" and "applied" epistemologies was revealed in an earlier stage of this study (Ronen-Fuhrmann et al., 2008); at the beginning of the semester, when engaged in theoretical discourse, students tended to advocate socio-constructivists paradigms, whereas when engaged in designing their technology-based modules, more than half of the students tended to neglect these ideas and apply more traditional approaches.

In order to quantitatively assess "low" and "high" levels of epistemology, we relied on studies that describe an instructionist view of learning as equivalent to naïve epistemology, and a socio-constructivist view, as analogous to sophisticated epistemology (Maor & Taylor, 1995; Schommer, 1990). Inductive analysis of our data indicated that there are three main dimensions in students' sayings and artifacts in the design process, which express their epistemologies: (a) learner activity, (b) collaboration, and (c) content accessibility. We used these dimensions to develop the epistemology rubric (Table 3)

Table 3: Epistemology rubric.

| Dimension | Low | Medium | High |
|---|---|---|---|
| **Learner activity**<br>The degree to which students expressed ideas that support active engagement of learners within a technology-based learning environment. | Passive: e.g. learner reads or views information. | E.g. learner clicks on links. | Active: e.g. learner manipulates variables |
| **Collaboration**<br>The degree to which the students supported using technology in ways that enable learners to learn from each other | Individual learning | Group work is not supported by technology | Collaboration is intrinsic to the activity |
| **Content accessibility**<br>The degree to which students expressed views that support making the contents of a learning environment accessible to learners. | No effort to connect contents to student world | Motivational aspects are extrinsic to activities | Motivational aspects are intrinsic to activities |

We would like to note that we used the epistemology rubric to assess the level of epistemology expressed in the artifacts created by *groups* of students rather than by individuals. Although in many studies epistemology is attributed to the individual, our rationale for studying group epistemology is based on the work of researchers such as Fuller (1987) and Goldman (2008), who claimed that there is a strong social aspect to epistemology, which cannot be overlooked; the individual's way of thinking is considerably influenced by the ideas expressed by the group of people he or she is interacting with.

## Data Analysis Procedure

To assess the design artifacts of the 14 case-study groups using the two rubrics, we first created digital portfolios for each group, which included all in-progress and final documents created by the group, all of the individual students' posts in the course's discussion forums, transcriptions of interviews (if there were any for students in the group), and their reflective essays. Each portfolio was organized by the stage of the design process (Analysis, Brainstorm, Build flow etc.). To refine initial versions of both rubrics, we took one example case-study, and had four researchers—two external researchers and the two authors of this paper—assess the degree of concretization and epistemology of the group at each stage of the design process. Initially, about 60% agreement was reached. Following several cycles of refinement of the rubric, in which more case studies were assesses, we reached a degree of about 90% agreement between researchers using the current version of the rubrics. In this manner seven of the portfolios (50% of the cases) were assessed. The rest of the case-studies were then assessed by the authors of this paper together (without comparison of individual assessments).

## Combining the Two Rubrics to Map Findings

Initial analysis of the findings showed that using each of the rubrics described above, we can distinguish between two patterns of learning. Using the concretization rubric, we found that one pattern was represented by groups who had difficulties in concretizing their design ideas. The concretization level of their artifacts at various stages of the design process was lower than the level required at that stage (see Table 2). On the other hand there were groups whose pattern did not show any difficulty with the concretization and were sometimes even ahead of the required level in the design process. This enabled us to refer to the dichotomy: *Low versus High* pace of concretization skills acquisition. Using the epistemology rubric, we were able to clearly distinguish between one pattern, in which groups of students showed a gap at beginning stages of the semester, as described above, versus another pattern of those who showed a coherent epistemology throughout the semester. This enabled us to refer to the dichotomy: *Coherent versus Non-coherent* pattern of group epistemology. Using these two dichotomies, we developed a four-quadrant matrix (Figure 2) to map our findings regarding the relations between concretization and epistemology.

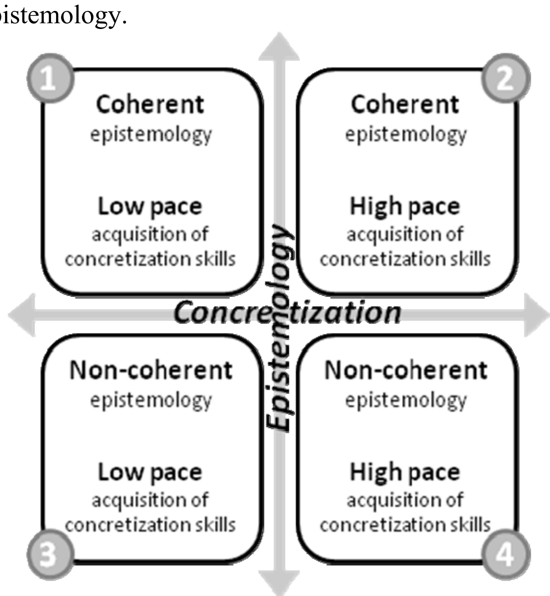



Figure 2. The four-quadrant concretization/epistemology matrix.

## Outcomes

Following an in-depth analysis of each of the 14 case-studies, in which we used both the concretization and the epistemology rubrics using all the data sources, we were able to map the cases into the four-quadrant matrix. The results of this mappying are represented in table 4. As can be seen from the the table, two of the cases were mapped in quadrant 1, another two in quadrant 2, three more in quadrant 3, and 7 cases—more than half of the students—were mapped in quadrant 4. Additionally, the anaysis of each of the cases' patterns of learning revealled that groups that were classified as belonging to quadrant 4 significantly reduced their episteomogical gaps throughout the semester, whereas groups that belonged to quadrant 3 only did so to a small extent. We argue that the high pace of their acquisition of concretization skills was an important factor in enabling groups in quadrant 4 to reduce their epistemological gaps. To support this claim, which we further discuss in the final section of this paper, we first describe in detail one case-study representing and illustrating the learning processes of groups that were classified as belonging to quadrant 4.

Table 4: Mapping of the 14 case-studies using the concretization/epistemology matrix.

| ***Quadrant*** | ***# of Groups*** | ***# of Students*** | ***Pattern*** |
|---|---|---|---|
| **1** | 2 | 5 | Coherent epistemology throughout the semester<br>Low pace of acquiring concretization skills |
| **2** | 2 | 4 | Coherent epistemology throughout the semester and<br>High pace of acquiring concretization skills |
| **3** | 3 | 6 | Non-coherent epistemology at beginning of semester<br>***Minor reducing of epistemological gap***<br>Low pace of acquiring concretization skills |
| **4** | 7 | 18 | Non-coherent epistemology throughout the semester and<br>***Major reducing of epistemological gap***<br>High pace of acquiring concretization skills |
| **Total** | 14 | 33 | |

## Illustrating Learning Processes in Quadrant 4: The case of I,S&E

I,S&E designed a technology-based module designated for high-school computer-science learners. Their module focused on recursive algorithms for scanning data-structure trees. One of the features they designed, from very early stages of the design process was an animation that demonstrates a certain algorithm for scanning a tree. Their (potential) learners were required to solve problems that utilize the demonstrated algorithm.

Analysis of the design artifacts they produced at various stages of the design process using the concretization rubric (left graph in Figure 3) indicates that this group's acquisition of concretization skills was of a high pace (high pace was defined as a slope that is higher or equals to 0.75, where each stage of the design course was numbered consecutively starting with "*Analysis=1*"). I,S&E had come up with the idea of the animation as early as the *Analysis* stage (in which they were still not required to suggest ideas for activities). They continued at a "normal", or "required" pace (see dotted line in Figure 3 - left graph) in the *Brainstorm Activities* and *Build Flow* stages. When required to design features, they were still struggling with their flow of activities, but they gradually progressed until they reached level four of concretization in their final mockup.

The analysis of IS&Es' learning process using the epistemology rubric (Figure 3, right side) revealed that at the beginning of the semester, in general discussions about educational technologies, each of these students expressed ideas that we classified as high level of sophistication with regards to epistemology (level 3 in each of the dimensions of the epistemology rubric). However, as can be seen in figure 3, there was a large drop at the *Analysis* stage, with respect to the *Learning Activity* and *Content Accessibly* dimensions, which continued with a drop of the *Collaboration* dimension at the *Build Flow* stage. These drops represent the gap described earlier, between "theoretical" and "applied" epistemologies. Specifically, when IS&E began to design their animation, it required learners only to passively watch the animation, and there was no attempt to make the contents more accessible. Collaborative aspects were minimal (a forum was designed for Q&A). Gradually, as this feature was revised following discussions with peers and instructors, and following the use of the Design Principles Database, this feature became a manipulable tool, which enabled learners to solve problems by exploring various ways to scan given trees, as well as their own trees. Our analysis of their final mockup, using the epistemology rubric was as follows: Learning Activity = 3 (learner manipulates variables); Content Accessibility = 2 (motivational aspects were eventually at an intermediate level); Collaboration = 3 (activities that required learners to solve problems in tasks created by their peers were designed). Thus - we interpreted their learning process as representing a major reduction of their groups' epistemological gap. The dotted line in left graph of Figure 3, which represents the average between the three dimensions, illustrates this reduction of the epistemological gap.

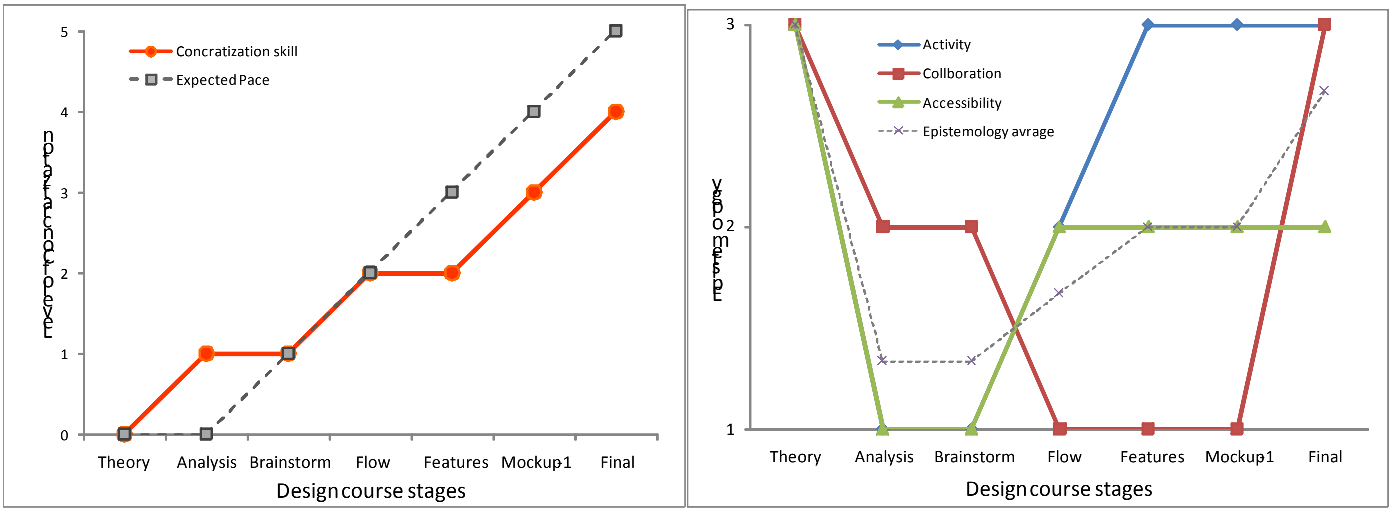



Figure 3: Analysis of the learning processes of I,S&E: Acquisition of concretization skills (left), and changes in group epistemology (right)

## Discussion

The mapping of the 14 case-studies using the four-quadrant concretization/epistemology matrix, revealed the important role of concretization in helping students to reduce epistemological gaps. The cases mapped in quadrants 1 and 3 illustrate that moving from general design ideas to actual activities was a difficult task for many of the students. Concretization of a high level, in the context of educational technology design, requires not only to design the details of certain features, but also to be able to describe how these features fit together in a coherent learning environment, and design different paths learners will be able to use in the environment. This requires a designer to deal with very high level of detail of the activities, and at the same time envision a flow of activities or scenarios that would enable a potential learner to acquire knowledge using the designed module.

Our findings illustrate that as students concretized their design ideas and represented them in sequences of activities, they exposed their pedagogical way of thinking to others. This enabled them to negotiate and reexamine their thinking with peers and instructors and to compare the design solutions they came up with, with those of experts. The exposure of ideas, induced by the concretization, brought to identification of gaps between students' views about how people learn, and pedagogical notions expressed in artifacts they designed at initial stages of the course (Ronen-Fuhrmann et al., 2007). Their views of learning, as stated in discussions, usually represented socio-constructivist approaches, while many of their initial design artifacts represented more naïve, transmissionist views of learning. As students' artifacts became more concrete, they also represented more advanced pedagogical views of learning. The gaps were reduced as a result of refinements students made throughout the design process. Following Wilensky's (1991) notion of concretization as representing the richness and connectedness of abstract concepts, we argue that a person who can verbally explain abstract pedagogical ideas, but has a difficulty to concretize them as features in a learning environment, lacks a deep and connected understanding of these ideas, and of the theoretical underpinnings behind them. To design concrete features that convey their ideas about learning, students in the current research were engaged in a process of connecting their ideas with design knowledge of peers, instructors, and experts. This process enabled them to develop a richer set of connections in their knowledge about learning theory, and thus develop a deeper understanding about learning.

Thus, in the context of educational technology design, we view concretization as: (a) an essential skill in the process of gaining design knowledge, and (b) a way to assist students to reflect and reduce gaps in their understanding about learning theory. This twofold notion of concretization is in agreement with the Learning By Design literature (e.g., Harel, 1991; Kafai, 2006; Kolodner et al., 2003; Papert, 1991), which shows that when learners design an artifact that explains a certain topic in a certain subject matter, they learn many things about design, but they also learn a whole lot about the topic they are designing for. In this study, the topic was not only the science or math embedded in the students' modules, but also the learning theory behind the instruction that they designed.

## References

Dick, W., Carey, L., & Carey, J. O. (2001). *The systematic design of instruction (5th ed.).* New York: Addison-Wesley Educational Publishers.

DiGiano, C., Goldman, S., & Chorost, M. (2009). *Educating learning technology designers: Guiding and inspiring creators of innovative educational tools.* N.Y.: Routledge

Fuller, S. (1987). On regulating what is known: A way to social epistemology. *Synthese, 73*, 145-183.

Goldman, A. (2008). Social epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2008 ed.(.

Harel, I. (1991). *Children designers: Interdisciplinary constructions for learning and knowing mathematics in a computer-rich school.* Norwood, NJ: Ablex Publishing.

Hoadley, C., & Kim, D. (2003). Learning, design and technology: The creation of a design studio for educational innovation. In A. P. d. Reis & P. Isaías (Eds.), *Proceedings of the IADIS international conference e-Society 2003* (pp. 510-519). Lisbon, Portugal: IADIS Press.

Kafai, Y. (2006). Playing and making games for learning. *Games and Culture, 1*(1), 36-40.

Kali, Y. (2006). Collaborative knowledge building using the Design Principles Database. *International Journal of Computer Support for Collaborative Learning, 1*(2), 187-201.

Kali, Y., & Linn, M. C. (2007). Technology-enhanced support strategies for inquiry learning. In J. M. Spector, M. D. Merrill, J. J. G. V. Merriënboer & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology (3rd Edition)* (pp. 445-490). Mahwah, NJ: Erlbaum.

Kali, Y ,.Linn, M. C., & Roseman, J. E. (2008). *Designing coherent science education: Implications for curriculum, instruction, and policy.* NY: Teachers College Press.

Kolodner, J. L., Crismond, D., Fasse, B., Gray, J., Holbrook, J., & Puntembakar, S. (2003). Putting a Student-Centered Learning by Design™ Curriculum into Practice: Lessons learned. *Journal of the Learning Sciences, 12*, 495 - 548.

Maor, D., & Taylor, P. C. (1995). Teacher epistemology and scientific inquiry in computerized classroom environments. *Journal of Research in Science Teaching, 32*(8), 839-854.

Papert, S. (1991). Situating constructionism. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 1–11). Norwood, NJ: Ablex.

Ronen-Fuhrmann, T., Kali, Y., & Hoadley, C. M. (2008). Helping education students understand learning through designing. *Educational Technology, 48*(2), 26-33.

Schommer, M. (1990). The effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*(3), 498-504.

Schon, D. (1983). *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Stake, R. E. (1994). Case studies. In N. K. Densin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 236-247). Thousand Oaks, CA: SAGE Publications.

Wilensky, U. (1 .(991Abstract meditations on the concrete and concrete implications for mathematics education. In I. Harel & S. Papert (Eds.), Constructionism. Norwood, MA: Norwood, MA: Ablex Publishing.

## Acknowledgements

This article is based on work supported by the National Science Foundation under grant number ESI-0334199. Any opinions, findings, and conclusions or recommendations expressed are ours and do not necessarily reflect the views of the National Science Foundation. We benefited from helpful conversations with the Technology-Enhanced Learning in Science group centered at the University of California, Berkeley, and with the Design Group at the Department of Education in Technology and Science at the Technion - Israel Institute of Technology. Finally, we thank all those who contributed from their design knowledge and experience to the Design Principles Database.

# Representational Scripting to Support Students' Online Problem-solving Performance

**Abstract:** This study investigated the effects of representational scripting on student learning while online collaboratively solving a complex problem. The premise here is that effective student interaction would be evoked when the problem-solving task is structured into part-tasks that are supported by providing part-task congruent representations (i.e., representational scripting). It was hypothesized that such an approach would lead to a more appropriate student interaction and as a consequence better problem-solving performance. In triads secondary education students worked on a case-based business-economics problem in four experimental conditions, namely one condition in which the groups received representations that were congruent for all three part-tasks and three conditions in which the groups received one of these representations for all three part-tasks. The results show that using representational scripting indeed leads to a more elaborated discussion about the content of the knowledge domain (i.e., concepts, solutions and relations) and to better problem-solving performance.

## Introduction

Collaboratively solving complex problems is often regarded as an effective pedagogical method that is beneficial for both group and individual learning. The premise underlying this approach is that through a dynamic process of eliciting one's own understanding of the content of the knowledge domain (i.e., concepts, principles and procedures) and discussing this with peers, students acquire new knowledge and skills and process them more deeply (e.g., Hmelo-Silver, Duncan, & Chinn, 2007). Unfortunately, putting students in groups and having them work together on a problem is not always beneficial for learning (e.g., Barron, 2003). Research on Computer Supported Collaborative Learning (CSCL) has shown that proper use of representational tools can beneficially affect externalizing, sharing and negotiating about the concepts, principles and procedures of the domain (e.g., Fisher, Bruhn, Gräsel, & Mandl, 2002). However, whereas these studies show promising results, other research questions how representational tools can best guide student interaction that is beneficial for learning (e.g., Suthers, 2006). Important here is that problem-solving tasks are usually composed of fundamentally different part-tasks, that each require the support of a different representational tool. Important here is that the guidance the tools are intended to provide is matched to the demands of the different part-tasks. Otherwise this will hinder learning (e.g., Van Bruggen, Boshuizen, & Kirschner, 2003). Recently, *scripting* has been advanced as a way to ensure the alignment between tool, tool use and learning goals in collaborative learning (e.g., Weinberger, Ertl, Fischer, & Mandl, 2005). Scripting the problem-solving process with representation tools sequences the problem-solving process and makes the different part-task demands explicit so that they can be foreseen with task-congruent content-related guidance by the representational tools. By doing so, part-task related activities beneficial for collaborative problem-solving can be evoked.

## Collaboratively Solving a Complex Problem

Collaboratively solving a complex problem is regarded to be a sequenced, phased process (i.e., problem orientation, problem solution, solution evaluation) where students actively engage in a process of sense-making in a knowledge domain, articulating and discussing multiple problem perspectives and problem-solving strategies (Ploetzner, Fehse, Kneser, & Spada, 1999; Van Bruggen, et al., 2003). Coping with the task demands of the different problem phases (i.e., part-tasks), requires students to interact in the *content space*; carry out part-task related activities such as discussing the concepts, principles, and procedures of the domain (Barron, 2003; Jonassen & Ionas, 2008). However, where expert problem-solvers experience no difficulties in carrying out these kinds of activities, students (i.e., non-experts) do. When solving problems, students rely primarily on surface features such as using objects referred to in the problem instead of the underlying principles of the knowledge domain, and employ weak problem-solving strategies such as working via a means-ends strategy towards a solution (Chi, 1997). Important here is that students lack a well developed understanding of the knowledge domain and as a consequence have problems creating and combining meaningful problem representations. This hinders students in effectively and efficiently carrying out their problem-solving task because the ease with which a problem can be solved often depends on the quality of the available problem representations (Chi; Jonassen & Ionas). To this end, it would be beneficial if suitable representations were provided and combined in a part-task appropriate manner.

## Representational Scripting

Integrating scripting with the availability of representational tools (i.e., representational scripting) structures the problem-solving process making it more efficient and effective. Scripting shapes the use of the representational tools and therefore also the epistemic and social processes of collaboration (Weinberger, et al., 2005) by sequencing and making the different part-task demands explicit so that they can be foreseen with task-congruent content-related guidance by the representational tools. Different representation tools provide different domain-specific content schemes (i.e., problem representation) representing different problem perspectives. Visualizing the knowledge domain through multiple external representations (ERs) influences student interaction by providing representational guidance (Suthers, 2006). The specific ontology (i.e., objects, relations, rules for combining them) of each ER offers a restricted view of the knowledge domain, guiding student interaction in a specific manner. Matching this representational guidance with the student interaction required to carry out a part-task evokes appropriate student interaction, leading to better problem-solving performance (see Table 1). To effectively do this, one must avoid or neutralize the difficulties encountered when combining multiple ERs, namely: translating from and coordinating between different ERs (Ainsworth, 2006), and incongruence between an ER and required part-task related activities (Van Bruggen, et al., 2003). This means that the representational guidance of the ER must be congruent (i.e., ontologically matched) to the part-task demands and activities of a specific problem phase. In this paper, the focus will be on guiding student interaction' when collaboratively solving a complex business economics problem.

Table 1: Congruence between external representation and part-task demands

| Problem phase | Part-task demands | ER | Representational guidance |
|---|---|---|---|
| Problem orientation | Determining core concepts and relating them to the problem | Conceptual (static) | Showing concepts and their interrelationship |
| Problem solution | Proposing multiple solutions to the problem | Causal (static) | Showing causal relation between the concepts and possible solutions |
| Solution evaluation | Determining suitability of the solutions and coming to a final solution to the problem | Simulation (dynamic) | Showing mathematical relation between the concepts and enabling manipulation of their value |

This study focuses on how the use of representational scripting affects both student part-task related interaction and problem-solving performance in a CSCL-environment. In four experimental conditions, student triads had to collaboratively solve a case-based problem in business-economics that was divided into three problem phases each coupled with a different ER. To study the effects of representational scripting, ERs were either matched or mismatched to the different problem phases; in other words they were either ontologically congruent or incongruent to the required part-task activities. In three mismatch conditions, groups received either a static ER (i.e., conceptual or causal ER) or a dynamic ER (i.e., simulation) that matched only one of the part-tasks. The scripting structured the problem-solving process in three phases, but only one of the ERs is available to the students for solving the problem, yielding a phase-match when the ER ontologically matched one of the three phases and a mismatch for the other two phases. In the fourth condition, groups received all three ERs in a phased order receiving the ER most suited to each problem phase. Here, thus, there was an ontological match between all three ERs and all three part-tasks. Due to the presumed match between ERs and part-tasks, student understanding and part-task related activities were expected to increase, allowing the students to reach better problem solutions. It was, therefore, hypothesized that students in the match condition *(H1)* carry out more part-task related activities and *(H2)* have better problem-solving performance.

## Method

### Participants

Participants were students from a business-economics class in a secondary education school in the Netherlands. The total sample consisted of 39 students (24 male, 15 female). The mean age of the students was 16.74 years (*SD* = .83, Min = 15, Max = 18). The students were randomly assigned to 13 triads divided between the four conditions; four triads in the conceptual and three triads in each of the other conditions (i.e., causal, simulation and match).

### Problem-solving task and materials

CSCL-environment: Virtual Collaborative Research Institute

Students worked in a CSCL-environment called Virtual Collaborative Research Institute (VCRI, see Figure 1), a groupware application for supporting the collaborative performance of problem-solving tasks and research projects (Broeken, Jaspers, & Erkens, 2006). For this study, five tools that are part of the VCRI were augmented with representational scripting. All tools, except the Notes tool, were shared among group members.

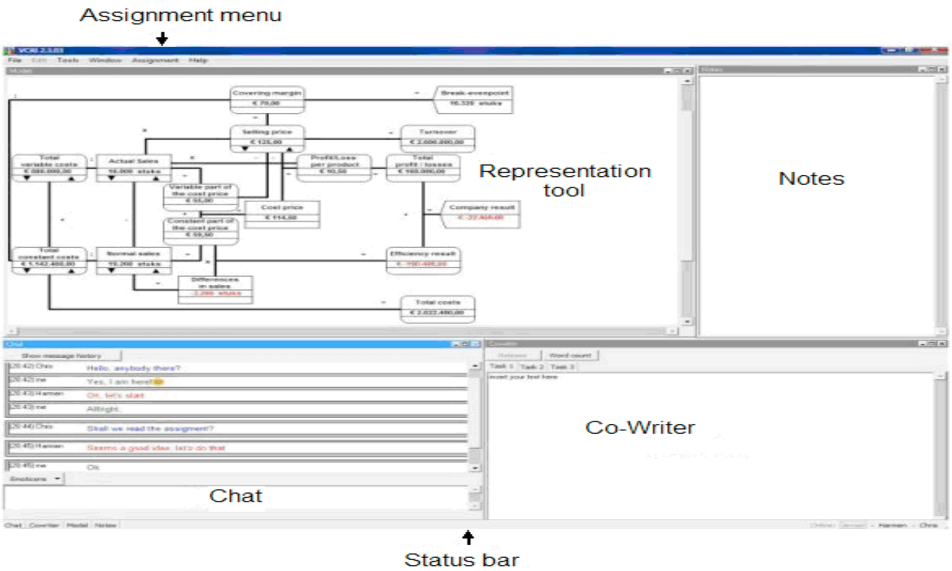



Figure 1. Screenshot of the VCRI-program.

The *chat tool* enables synchronous communication and supports students in externalizing and discussing their knowledge and ideas. The chat history is automatically stored and can be re-read by the students. In the *Assignment menu*, students can find the description of the problem-solving task / part-tasks. Besides this, additional information sources such as a definition list, formula list, and clues for solving the problem were also available here. The *Co-writer* is a shared text-processor where students can collaboratively formulate and revise their answers to the part-tasks. The *Notes tool* is an individual notepad that allows students to store information and structure their own knowledge and ideas before making them explicit. The *Status bar* is an awareness tool that displays which group members are logged into the system and which tool a group member is currently using. All students in all conditions had access to these tools and information sources. In other words, the different conditions were information equivalent and only differed in the way that the ERs are intended to guide the interaction.

## Problem-solving task and design representational scripting

All groups worked on a case-based problem in business-economics in which they had to advise an entrepreneur about changing the business strategy to increase profits (i.e., company result). To come up with an advice, students had to carry out three different part-tasks, namely (1) determine the main factors affecting the company's result and relate them to the problem, (2) determine how certain interventions affect company result, and (3) compare the effects of these interventions and formulate a final advice based on this comparison. Through scripting, the problem-solving process was structured into a problem orientation phase, problem solution phase, solution evaluation phase each focusing on one of the part-tasks. All groups were 'forced' to carry out all the part-tasks in a predefined order; they could only start with a new part-task after finishing an earlier part-task. When group members agreed that a part-task was completed, they had to 'close' that phase in the assignment menu. This 'opened' a new phase, which had two consequences for all groups, namely they (1) received a new part-task, and (2) had to enter their new answers in a different window of the Co-writer. All conditions received the part-tasks in the same order, but only groups in the match condition received a new ER.

The *problem orientation phase* focused on creating a global qualitative problem-representation by asking students to explain what they thought the problem was and to describe what the most important factors were that influenced the problem. During this phase, students received the conceptual ER (i.e., a static content scheme; see Figure 2) that made two aspects salient, namely the core concepts needed to carry out this part-task and how the core concepts are qualitatively interrelated. Students could see that 'company result' is affected by the 'total profit' and the 'efficiency result'. Such information should make it easier for them to create an overview of all relevant concepts, supporting them in finding multiple solutions to the problem in the following phase.

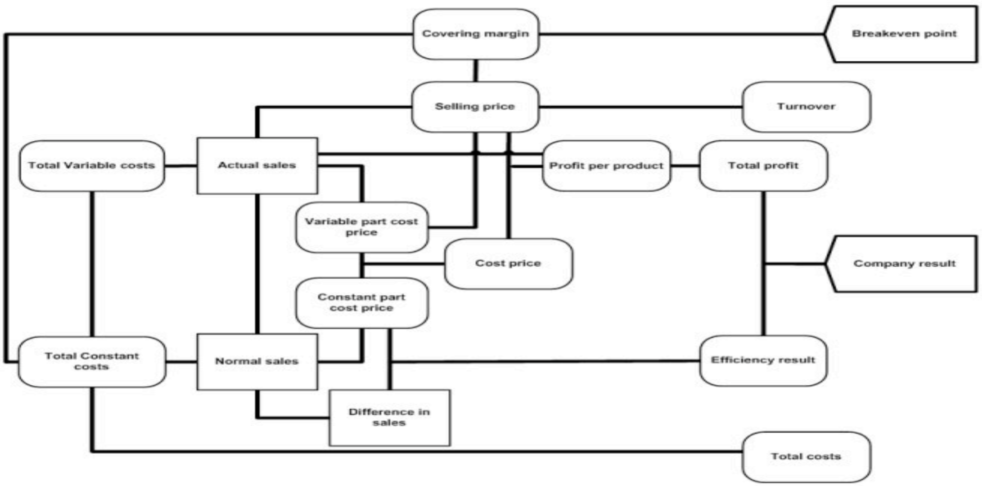



Figure 2. Conceptual ER.

The *problem solution* phase focused on creating a causal problem representation (i.e., explicating the underlying business-economics principles) by asking students to formulate several solutions to the problem. During this phase, students received the causal ER (i.e., a static content scheme; see Figure 3), in which the causal relationships - visible through the arrows showing direction of the relationship between the concepts - were specified. The causal ER also contributed to increasing students' qualitative understanding by providing the students with possible interventions (i.e., changes in the business strategy), each of which had a different effect on the company results. This should make it easier to effectively explore the solution space and should, in turn, support students in finding multiple solutions to the problem. Students could, for example, see that a PR-campaign affects 'actual sales that in turn affects 'total profit'.

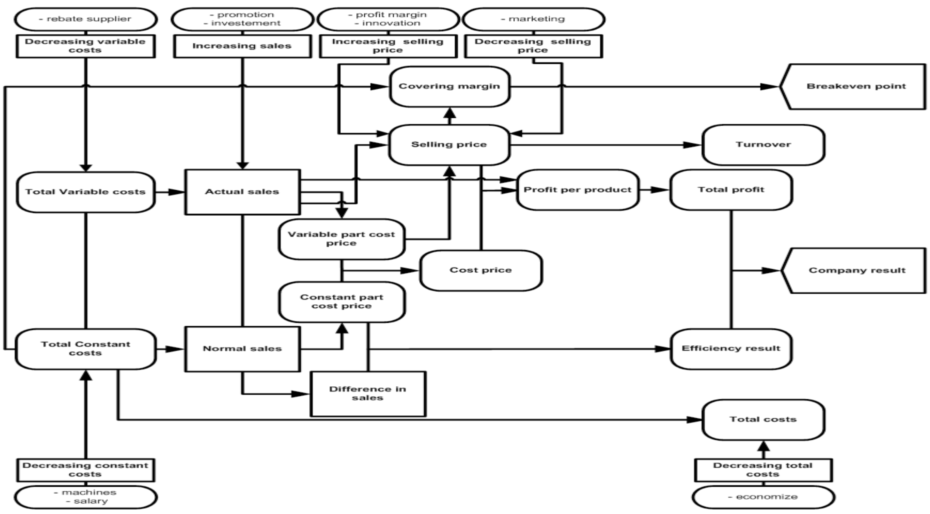



Figure 3. Causal ER.

Finally, the *solution evaluation phase* focused on increasing the students' understanding of the knowledge domain with the aid of a quantitative problem representation. Students were asked to determine the financial consequences of their proposed solutions and to formulate a final advice for the entrepreneur by negotiating the suitability of the different solutions with each other. During this phase, students received a simulation ER (i.e., a dynamic content scheme; see Figure 4) that enabled them to manipulate the values of the concepts by clicking on the arrows in the boxes. When the value of a certain concept was changed (i.e., increased or decreased), the simulation model automatically computed the values of all other concepts. This is meant to facilitate the determination and negotiation of the suitability of the different proposed solutions and reaching a final advice. Students could, for example, test how the PR-campaign affects the 'actual sales' and whether this in turn affects the 'total profit'. Only the simulation ER is capable of providing this kind of support, because the relationships between the concepts in this ER were specified as equations.

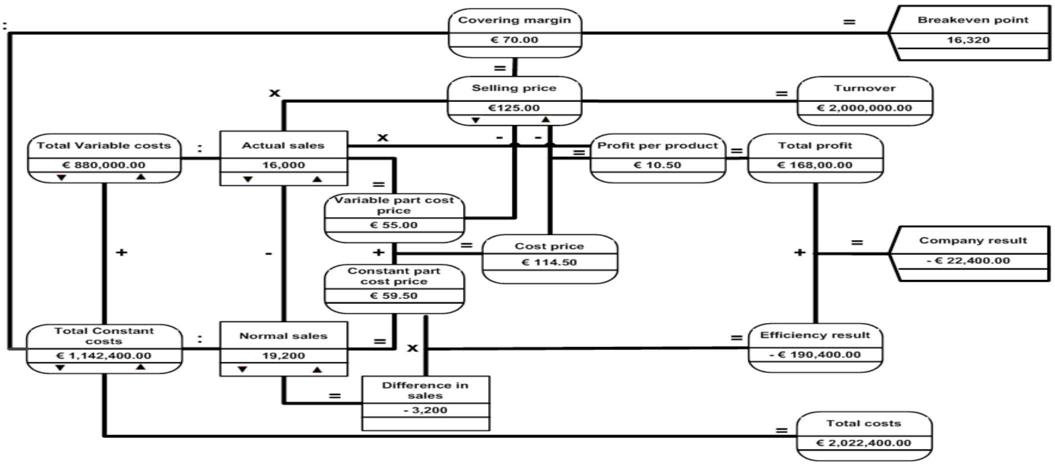



Figure 4. Simulation ER

## Procedure

All groups spent three, 70-minute lessons solving the problem during which each student worked on a separate computer in a computer classroom. Before the first lesson, students received an instruction about the CSCL-environment, the group composition, and the problem-solving task. The instruction made it clear that their group answer to the problem (i.e., problem-solving performance) would serve as a grade affecting their GPA. Students worked on the problem in the computer classroom, where all actions and answers to the part-tasks were logged.

## Measures

### Content-related student interaction

To examine the effect of condition data concerning student interaction was collected by logging the chat-utterances of the group members. The content of these chat-protocols is assumed to represent what students know and consider important for carrying out their problem-solving task (Chi, 1997). Using so called 'concordancers' software (e.g., MEPA, see Erkens & Jansen, 2008) minimizes the work associated with coding chat-protocols and maximizes coding reliability allowing the content of chat-protocols to be searched for the occurrence of characteristic words (i.e., key words) that led to the identification and coding of the dependent variables. This was done automatically with a MEPA-filter using 'if-then' decision rules containing different explicit references to a concept, solution or relation (e.g., name, synonyms) that were coded as representing that concept, solution or relation (see Table 2).

Table 2: Concepts, solutions and relations; coding and reliability MEPA-filter

| Categories | Subcategories | Discussion of | Reliability |
|---|---|---|---|
| Concepts | Sales | how many products are sold / have to produced | 90% |
| | Selling price | what the customer has to pay for the product | |
| | Costs | what the overall costs of the company are | |
| | Turnover | what the total income of the company is | |
| | Company result | whether it is profitable to run the company | |
| Solutions | Changing costs | how the overall costs can be decreased | 84% |
| | Changing turnover | how the turnover can be increased | |
| | Changing both | the combining of the other two solutions | |
| Relations | Conceptual | the definition / meaning of a concept / solution | 80% |
| | Causal | the causal relationship within / between concepts / solutions | |
| | Mathematical | the quantitative relationships within / between concepts | |

### Problem-solving performance

To measure the effect of condition on problem-solving performance an assessment form for all criteria of the problem-solving task was developed (see Table 3). The 41 items were coded as; 0, 1 or 2, whereby a '2' was coded when the answer given was of high quality. All groups could, thus, achieve a maximum score of 82 points (41 * 2 points).

Table 3: Problem solving performance; items and reliability

| Criteria | Description | Items | α |
|---|---|---|---|
| Suitability | Whether the groups' answers were suited to the different part-tasks | 9 | .81 |
| Elaboration | Number of different business-economics concepts or financial consequences incorporated in the answers to the different part-tasks | 9 | .56 |
| Justification | Whether the groups justified their answers to the different part-tasks | 9 | .71 |
| Correctness | Whether the groups used the business-economics concepts and their interrelationships correctly in their answers to the different part-tasks | 9 | .68 |
| Continuity | Whether the groups made proper use of the answers from a prior problem phase | 2 | .67 |
| Quality advice | Whether the groups gave a proper final advice<br>- Number of business-economics concepts incorporated in the advice<br>- Number of financial consequences incorporated in the advice<br>- Whether the final answer conformed to the guidelines provided | 3 | .76 |
| Total score | Overall score on the problem-solving performance | 41 | .92 |

### Data analyses

When conducting studies on CSCL, group members mutually influence each other (i.e., behave more or less in the same way) leading to non-independence of measurement (Kenny, Kashy, & Cook, 2006). This is problematic because many statistical techniques (e.g., *t*-test) assume score independence and such a violation compromises the interpretation of the analyses (e.g., *t*-value, standard error, see Kenny, et al.). The non-independence was determined here by computing the intraclass correlation coefficient and its significance (Kenny, et al.), for all dependent variables concerning student interaction. This resulted in non-independence ($\alpha < .05$) for all tests, justifying *Multilevel analysis* (MLA) for analyzing these data. MLA entails comparing the deviance of an empty model and a model with one or more predictor variables to compute a possible decrease in deviance. The latter model is considered a better model when there is a significant decrease in deviance in comparison to the empty model (tested with a $\chi^2$-test). All reported $\chi^2$-values were significant ($\alpha < .05$) and, therefore, the estimated parameter of the predictor variables (i.e., effects of condition) were tested for significance. Due to the detection of outliers, the utterances of some students were deleted from the MLAs. One-way MANOVA was used for answering the second research question. Since there were specific directions of the results expected (see hypotheses) all analyses are one-sided.

## Results

### Concepts, solutions and relations

MLAs revealed that experimental condition was a significant predictor for the number and kinds of concepts, solutions and relations that students discussed (see Table 4). The results show a main effect for *concepts*, students in the match condition discussed significantly more concepts in comparison to students in three non-matched conditions ($\beta = 9.08$, $p = .00$). When comparing the match condition to the other conditions separately, it appeared that students in the conceptual condition discussed concepts as 'sales' ($\beta = 2.03$, $p = .03$), 'turnover' ($\beta = 1.80$, $p = .00$) and 'company result' ($\beta = 2.55$, $p = .02$) less frequently than students in the match condition. The same kind of results were obtained when comparing students in the match condition to students in the simulation condition; 'sales' ($\beta = 2.00$, $p = .03$), 'turnover' ($\beta = 2.28$, $p = .00$) and 'company result' ($\beta = 3.19$, $p = .02$). Second, a main effect for *solutions* was found. Students in the match condition discussed significantly more solutions in comparison to students in the non-matched conditions ($\beta = 6.33$, $p = .00$). When comparing the match condition to the other conditions for the different kinds of solutions, students in the match condition discussed the solution aimed at increasing the company's turnover more often than students in the non-matched conditions ($\beta = 3.42$, $p = .03$). This effect was also significant between all conditions. Finally, Table 4 shows a main effect for *relationships*, students in the match condition discussed significantly more and different kinds of relationships than students in the non-matched conditions ($\beta = 11.68$, $p = .00$). When comparing the match condition to the other conditions separately, it appeared that (1) students in the conceptual ($\beta = 2.85$, $p = .03$) and the simulation condition ($\beta = 3.67$, $p = .02$) discussed significantly fewer conceptual relationships, (2) students in the conceptual condition discussed significantly fewer causal relationships (β = 6.41, *p* = .01), and (3), students in the conceptual (β = 2.50, *p* = .01) and the simulation (β = 2.68, *p* = .00) condition discussed fewer mathematical relationships.

Table 4: Multilevel Analyses for Effects concerning Students' Discussion of Concepts, Solutions and Relations

| | Conceptual condition (*n* = 10) | Causal condition (*n* = 10) | Simulation condition (*n* = 10) | Match condition (*n* = 6) | Effects match condition (*N* = 36) | | |
|---|---|---|---|---|---|---|---|
| | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* | $\chi^2(3)$ | β | *SE* |
| *Concepts* | 8.36 (5.20) | 10.88 (8.64) | 4.50 (4.70) | 26.50 (14.57) | 25.37 | 9.08** | 2.00 |
| Sales | 1.09 (1.30) - | 2.00 (2.14) | 1.20 (1.14) - | 5.17 (6.40) + | 11.53 | 2.03** | 0.72 |
| Selling price | 0.27 (0.65) | 1.00 (1.85) | 0.20 (0.63) | 2.33 (2.50) | 5.67 | 1.03 | 0.54 |
| Costs | 3.18 (3.09) | 3.50 (2.88) | 1.50 (2.17) | 6.50 (3.27) | 13.18 | 1.67 | 0.70 |
| Turnover | 1.09 (1.45) - | 1.25 (1.39) | 0.10 (0.36) - | 4.67 (3.08) + | 16.51 | 1.79* | 0.40 |
| Company result | 2.73 (2.15) - | 3.12 (3.14) | 1.50 (1.78) - | 7.83 (3.66) + | 14.06 | 2.55* | 0.81 |
| *Solutions* | 6.00 (4.65) | 7.12 (5.69) | 2.80 (3.19) | 18.67 (7.74) | 23.29 | 6.33** | 1.45 |
| Changing costs | 3.09 (3.53) | 3.75 (3.50) | 1.50 (2.12) | 6.33 (3.56) | 10.87 | 1.62* | 0.81 |
| Changing turnover | 2.18 (1.94) - | 2.88 (2.75) - | 0.90 (1.11) - | 10.33 (4.89) + | 23.22 | 4.08** | 0.70 |
| Changing both | 0.73 (1.27) | 0.50 (0.75) | 0.40 (0.52) | 2.00 (1.79) | 3.88 | 0.64* | 0.32 |
| *Relations* | 11.73 (5.31) | 18.37 (12.36) | 8.10 (3.81) | 35.17 (15.38) | 39.48 | 11.68** | 2.98 |
| Conceptual | 4.00 (2.97) - | 4.00 (5.16) | 2.30 (2.50) - | 9.67 (5.24) + | 13.71 | 2.86* | 1.25 |
| Causal | 6.91 (3.89) - | 2.38 (9.15) | 5.30 (3.40) | 19.67 (5.24) + | 19.76 | 6.29* | 1.93 |
| Mathematical | 0.82 (1.33) - | 2.00 (3.07) | 0.50 (0.85) - | 5.83 (5.42) + | 15.39 | 2.51* | 0.68 |

Notes. * $p < .05$; ** $p < .01$; if match condition significantly > a mismatch condition than the match condition is indicated with a + and the mismatch condition with a -

## Problem-solving performance

One-way MANOVA on the total score of the problem-solving performance showed a significant difference for condition ($F(3, 9) = 1.99$, $p = .04$; Pillai's Trace = 2.00; partial *eta* squared = .67). Bonferroni post hoc analyses showed that groups in the match condition indeed scored significantly higher than groups in both the conceptual ($p = .02$; $d = 2.28$) and the simulation condition ($p = .05$; $d = 1.90$). Table 5 shows the mean scores and standard deviations of the scores on the problem-solving performance. When the results for the dependent variables were considered separately, using one-way ANOVAs with Bonferroni post hoc analyses, condition effects were found for 'suitability' ($F(3, 9) = 4.49$, $p = .02$), 'elaboration' ($F(3, 9) = 3.13$, $p = .04$) and 'correctness' ($F(3, 9) = 4.25$, $p = .02$). The mean scores indicate that there were several significant differences between conditions. First, groups in the match condition scored significantly higher on 'suitability' than groups in both the conceptual condition ($p = .03$; $d = 3.61$) and the simulation condition ($p = .05$; $d = 3.28$). Second, groups in the match condition scored significantly higher on 'elaboration' than groups in the conceptual condition ($p = .04$; $d = 1.57$).

Table 5: One-way Multivariate Analysis concerning Group Problem-solving Performance

| Criteria | Condition | | | |
|---|---|---|---|---|
| | Conceptual condition (*n* = 4) | Causal condition (*n* = 3) | Simulation condition (*n* = 3) | Match condition (n = 3) |
| | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* |
| Suitability* | 10.75 (1.50) - | 13.67 (1.52) | 11.33 (4.16) - | 17.00 (1.73) + |
| Elaboration* | 3.75 (2.06) - | 6.67 (2.08) | 6.00 (2.00) | 9.67 (3.78) + |
| Justification | 3.25 (2.06) | 4.67 (3.06) | 3.00 (3.00) | 8.00 (4.00) |
| Correctness* | 4.50 (1.29) - | 6.33 (3.77) | 5.33 (0.58) - | 10.67 (2.89) + |
| Continuity | 2.00 (1.41) | 2.00 (1.00) | 2.00 (1.73) | 3.67 (0.58) |
| Quality advice | 2.50 (0.58) | 3.67 (0.58) | 3.33 (2.52) | 3.33 (1.53) |
| Total score* | 26.75 (4.17) - | 37.00 (7.00) | 31.00 (12.00) - | 52.33 (11.24) + |

Notes. * $p < .05$; ** $p < .01$; if match condition significantly > a mismatch condition than the match condition is indicated with a + and the mismatch condition with a -

# Discussion

As is the case with many other researchers, the present study stresses the importance of using representational tools (Fischer, et al., 2002) and of employing scripting (Weinberger, et al., 2005) to guide student interaction and problem-solving performance. However, instead of using them separately, this study combined the advantages of using multiple representations and scripting (i.e., representational scripting). The representational scripting structured the problem-solving process by sequencing and making the part-tasks explicit so that they could be foreseen with ontologically congruent content-related guidance in the representational tools. It was hypothesized that this would evoke more suited part-task related student interaction and, as a consequence, better problem-solving performance than not receiving it. Although based on 13 triads, the results of our study confirmed that the problem-solving process for these groups was more efficient and effective. Specifically, groups in the match condition had more elaborated discussions of the content of the knowledge domain, gave better answers to the part-tasks and a came up with better final solutions to the problem than groups in the non-matched conditions. Although the results seem very promising student interaction and problem-solving performance of students in the causal condition was very similar to what was found in the match condition. Students in both conditions received the causal ER, that showed all relevant concepts, solutions and their causal interrelationships, providing the students multiple qualitative perspectives on the knowledge domain. Providing ERs that foster causal reasoning seems, therefore, beneficial for collaborative problem-solving (Jonassen & Ionas, 2008). Combining the causal ER with both other ERs might also hinder problem-solving when students experience difficulties integrating the different ERs. When students do not know how to combine multiple ERs, they might choose to stick with the familiar one and make no attempt to integrate the different ERs (Ainsworth, 2006). Future research should be aimed at determining whether the results can be generalized to the same and other kinds of problem-solving tasks. Its seems also interesting to study the effects of the collaborative construction of external representations. Such an approach might lead to more insight into how students use and combine the concepts, solutions and relations within and between different ERs.

## References


Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16,* 183–198.

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12,* 307–359.

Broeken, M., Jaspers, J., & Erkens, G. (2006). Virtual Collaborative Research Institute (VCRI). Version 2.3. Utrecht, The Netherlands: Utrecht University.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6,* 271–315.

Erkens, G., & Janssen, J. (2008). Automatic coding of online collaboration protocols. *International Journal of Computer-Supported Collaborative Learning, 3,* 447–470.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12,* 213–232.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42,* 99–107.

Jonassen, D. H., & Ionas, I. G. (2008). Designing effective support for causal reasoning. *Educational Technology Research and Development, 56,* 287–308.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York/London: The Guilford Press.

Kirschner, P. A., Beers, P. J., Boshuizen, H. P. A., & Gijselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior*, *24*, 403–420.

Ploetzner, R., Fehse, E., Kneser, C., & Spada, H. (1999). Learning to relate qualitative and quantitative problem representations in a model-based setting for collaborative problem solving. *Journal of the Learning Sciences, 8,* 177–214.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning,1,* 315–337.

Van Bruggen, J. M., Boshuizen, H. P. A., & Kirschner, P. A. (2003). A cognitive framework for cooperative problem solving with argument visualization. In P. A. Kirschner, S. J. Buckingham-Shum, & C. S. Carr (Eds.), *Visualizing Argumentation: Software tools for collaborative and educational sense-making.* (pp. 25-47). London: Springer. Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*, 1–30.

# Extending Students' learning Spaces: Technology-Supported Seamless Learning

Wenli Chen, Peter Seow, Hyo-Jeong So, Yancy Toh, & Chee-Kit Looi
National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616
Email: wenli.chen@noe.edu.sg, peter.seow@nie.edu.sg, hyojeong.so@nie.edu.sg, yancy.toh@nie.edu.sg, cheekit.looi@nie.edu.sg

**Abstract:** Learning is interweaved into and across students' everyday life activities. Technology that is used to support learning should be integrated with everyday life in the same way that learning occurs in everyday life: seamlessly. Mobile technologies, with their reduced size and ease of use, provide the potential to extend students' learning spaces and enrich the learning experiences in their daily lives where they move between locations, switch from one topic or context to another, and interact with different social groups. This paper proposes mobile technology-supported seamless learning and presents learning scenarios from our research to illustrate how learning occurs seamlessly across time and places mediated by mobile devices.

## Introduction

Learning is interweaved into and across students' everyday life activities such as watching videos, browsing Internet, reading book, playing sports, talking to people, and shopping. These activities can be resources and contexts for learning which should be seen not as a shell that surrounds the learner at a given time and location, but as a dynamic entity, constructed by the interactions between learners and their environment (Sharples, Taylor, & Vavoula, 2007). Looi, Seow, Zhang, So, et al (2009) state that the learning space is no longer defined by the 'class' constrained by scheduled class hours or specific locations, but by 'learning' across spaces as they harness ideas and learning resources gained in one location or context, and apply or develop them in another. They learn across time, by revisiting knowledge that was gained earlier in a different context. They learn by moving from topic to topic, managing a range of personal learning projects, rather than following a single curriculum (Sharples, Taylor, & Vavoula, 2007). They learn across social groups, by co-constructing knowledge with another student, a small group, or a large online community, with the possible involvement of teachers, friends, relatives, experts and members of other supportive communities.

Technology that is used to support learning should be integrated with everyday life in the same way that learning occurs in everyday life: seamlessly. Mobile technologies offer the potential for a new phase in the evolution of technology-enhanced learning, marked by a continuity of the learning experiences across different contexts. Chan, Rochelle, Kinshuk and Sharples et al. (2006) use the term "seamless learning" to describe the situations where students can learn whenever they want to in a variety of scenarios and that they can switch from one scenario to another easily and quickly using their personal mobile device as a mediator.

While there is growing recognition that learning can take place anywhere and anytime beyond the walls of classrooms, until recently learning across spatial and temporal spaces has remained an under-researched area. We are conducting a 3-year project designing and investigating a seamless learning environment to bridge Primary (Elementary) students' different learning spaces by equipping them with personal mobile devices. In this paper, we explore how to use mobile technology to support seamless learning and present learning scenarios from our research to illustrate how students use mobile technologies to extend their learning spaces.

## Seamless Learning Supported by Mobile Technology

The basic premise of seamless learning is that it is not feasible to equip students with all the knowledge and skills they need to have for lifelong learning based on snapshots of episodic time frame, location or scenario. Therefore, students will need to continually enhance their knowledge and skills, in order to address immediate problems and to participate in a process of continuous learning. We consider student learning as moving beyond the acquisition of content knowledge to developing the capacity to learn seamlessly.

As the use of mobile technology such as the Smartphone is becoming more pervasive in our lives, we envision learning to occur when a learner interacts with others and the environment across time and location seamlessly through the use of the technology as a mediating tool. Moving away from the Cartesian perspective where the learner is separated from the world, we view the learner as an active participant in the world. Learning occurs through the rich social interactions with others and interactions with environment as shown in Figure 1. Learning is developmental and occurs through the interactions a learner may experience at different points of time. Technology is a nexus connecting the learning experiences of the learner.

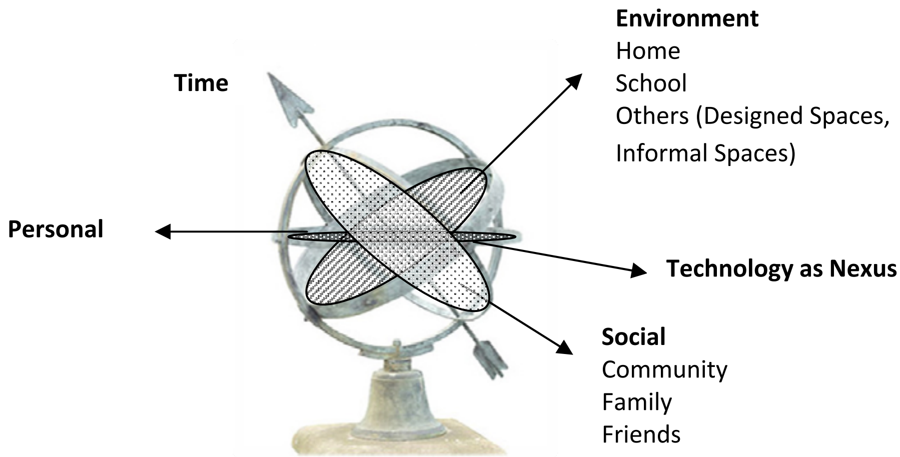



Figure 1. Technology supported seamless learning

Stroup and Petrosino (2003) categorize educational technology into two types: vertical technology and horizontal technology - vertical technology is mainly for teachers' needs in a confined setting whereas horizontal technology is used to meet students' personal needs across multiple physical contexts. Many of today's educational technologies are used as vertical technologies in the classroom. Seamless learning requires horizontal technologies that can meet students' personal needs. Mobile technology is exactly a horizontal technology because of its affordances. The characteristics of mobile technologies are well suited to support seamless learning (Chen, Seow, So, Toh, & Looi, in press). The small size and light weight of mobile devices mean that they can be taken to different sites or moved around within a site, so that they can be available wherever the student needs to learn. The use of the mobile device becomes a routine practice and is assimilated into everyday life experiences. Equipping with mobile technologies, students can learn across time and location. The learning experience supported by mobile technology is unobstrusive -it enables students' data exchange, communication and collaboration with teachers, experts, friends and family members etc. As the mobile device can be turned on and off instantaneously, students can use it whenever they need to, enabling them to make rapid connections between ideas and observations. The mobile device can both gather and respond to information specific to the current location, environment and time. This help student learn in real context.

Besides, mobile technology accommodates versatile learning activities. Students can type, draw, take photos, do audio or video recording, supporting their multi-modal expressions. The artifacts created by students can be saved. The student's personal accumulation of resources and knowledge is persistent and can be made immediately accessible by others. The use of the device is adapted to the learner's evolving abilities, skills, knowledge and learning styles and designed to support individualized learning, rather than general office work.

Some of the requirements of technology to support seamless learning can be satisfied by traditional tools such as notebooks, pencils, textbooks, timetables, and diaries etc. However, mobile technologies can supplement these by offering learners the opportunity to manage their learning over time, to engage in collaboration, and to relate information to situated problems. In sum, it is a *learning hub* for the students to engage in several learning activities across time and location by using the same devices.

In our research, each student is equipped with a HTC Tytn II Windows Mobile Smartphone that came with photo-taking function, stylus pen, keyboard, 3G-enabled Internet surfing data plan and educational applications. The students used the Smartphones in class and were allowed to bring home the devices. With the 3G-enabled Smartphone, they have instant access to the Internet inside and outside school at anytime. The students were trained in the basic use of the Smartphone such as the camera, internet connectivity, web browsing and GoKnow Mobile Learning Environment (MLE). GoKnow's MLE is a suite of educational mobile applications comprising Sketchy for creating animated drawings, PicoMap for creating concept maps, iKWL for creating KWL, and MyProjects for students to carry out their learning activities.

## Research Design

Our research is guided by the following research questions: 1) How does the appropriate use of mobile technologies mediate planned and emergent learning inside and outside school? 2) How do students learn seamlessly across the boundaries of learning spaces? and 3) How do we design for learning to occur seamlessly across the boundaries? To answer our questions, we studied an experimental Primary (Elementary) Grade 3 (P3, with nine-year old students) class comprising 39 students and worked closely with the class form and science teacher. The students have been using the HTC Smartphones since March 2009.

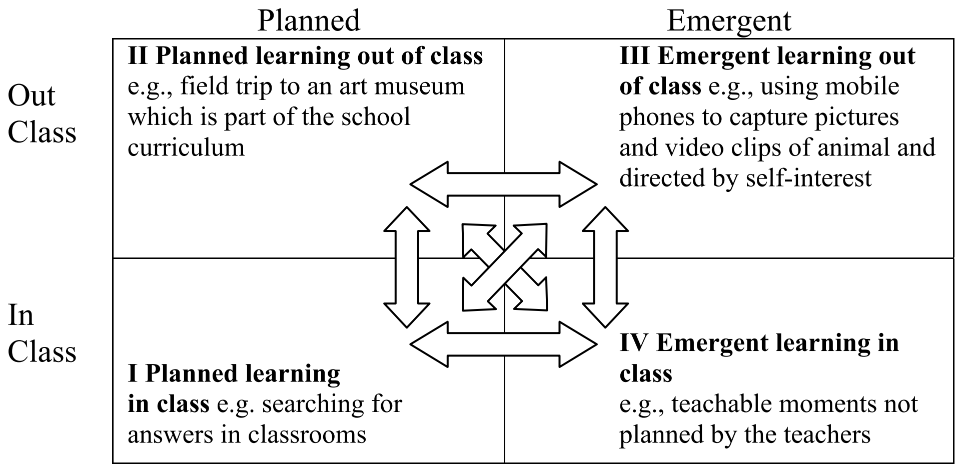



Figure 2: Matrix of students learning spaces (adapted from So, Kim, & Looi, 2008)

Our research attempts to design and investigate a seamless learning environment that allows students to learn at any time and at any location, and provides children with multiple ways of learning throughout the day. This is consistent with what Pea (2009) argues: "we need to treat the activities and life experiences of the learners throughout the day as our units of learning design, description and explanation". The matrix above (Figure 2) shows the students' learning spaces from two dimensions – 1) in class vs. out of class and 2) planned learning (planned by teachers) vs. emergent learning (not planned by teachers, but occurring unexpectedly driven by student self-interest/motivation).

### "Mobilized" Lesson Design

For the planned learning activities, we designed a series of "mobilized" lessons in class (Zhang, Wong, Seow, Chen, & Looi, 2009) to help students learn curriculum subjects better. Instead of relying on the textbooks for instruction, we created instructional tasks for the students to search for information related to the curriculum and scaffold their learning. The instructional tasks took advantage of the affordances of the Internet-enabled Smartphones and the GoKnow's MLE. For example, we used iKWL to draw out students' prior understanding and encourage inquiry about the topic. Students use the Internet-enabled Smartphones to search for resources to answer their queries. Many "mobilized lessons" have an "out-of-classroom" element where the students interact with the environment outside the classroom and engage in activities in their everyday lives. For example, we designed an English mobilized lesson on prepositions (Looi, Wong, So, Seow, et al, 2009) to support students' multiple learning paths. In this lesson, students used the Sketchy application on their mobile devices to represent their understanding of the prepositions, and took pictures outside the classroom to demonstrate their usage in an authentic context.

### Data Collection

A design ethnography approach (Goldman, 2004) is integrated into our research to observe how students are engaged in seamless learning when they interact with their handheld devices, peers, teachers and other people in their learning community (e.g., Barron, 2006; Squire & Klopfer, 2007). Over the past 6 months, we have followed this P3 class closely to study how they use the Smartphone in and outside of the classroom. Data was obtained through extensive observations in the school. The sources of data included classroom observations, informal conversation with students during their recess breaks, conversations with the teachers, and professional development sessions with teachers as we designed and reviewed the lessons. Our researchers spent time observing and interacting with the class every week for about 6 months. They observed how the students used the Smartphones to mediate learning in class, how the Smartphones facilitate interactions and learning with others, and the emergent use of the Smartphone.

We surveyed the students before the introduction of the Smartphone and again six months after they have used the device. In the survey, the students reported on their usage, experience and attitudes towards the use of the Smartphone for learning in and out of school. We conducted structured interviews with the students prior to their use of the devices.

In a seamless learning environment, students are 'on the move' across different modes of space and time. In this project, students have 24-7 access of the mobile device so that they can use the tool to engage in a wide range of activities in and out of the classroom. We have developed and installed a log software program to capture all the data stored in students' HTC phone. To understand more about the nature of students' learning outside school, we selected six students from a diverse set of criteria (academic performance, technology competency, socio-economic status, social skills and richness of informal learning spaces) for in-depth study. For these students, we conducted monthly visits to their home where we interviewed them and their family members. Through the visits, we hope to understand the informal learning environment of the student at home and outside school. Each family was given a digital camera to capture moments that reflect the student's use of the Smartphones during family activities.

## Findings

In this section, we describe a myriad of seamless learning scenarios based on our preliminary findings of the data collected from the research. The scenarios describe how technology facilitates learning in and across the boundaries presented in our framework in Figure 3. We present survey results on the 39 students' experiences with and attitudes towards the Smartphone and its role in learning 6 months after the introduction of HTC phones (September 2009) and "mobilized lesson. We also compare students' attitudes toward using Smartphone for their learning in and out of classroom before (February 2009) and half year after the introduction of HTC phones and "mobilized lesson".

### Seamless learning Scenarios

#### Planned Learning in the classroom – Planned learning outside class

In a lesson on digestive systems, the students learnt about how a digestive system functions as a system. They had to understand how different parts of the human digestive system worked together to digest food in the body. Planned classroom activities included viewing of videos on students' Smartphones and creating animation on the synergistic processes of the digestive parts. The process of creating the animation enabled students to grasp why and how the seemingly disparate parts actually work in tandem as a system. For the planned home activity, the teacher wrote a letter to the parents of the students to invite them to participate in the activity. Students were required to teach their parents on what they know about the digestive system. After the parents had learnt about the digestive system from the child, they were asked to share and recount the parts and functions of the digestive system. The students then used the Smartphone to video or audio record what the parents have shared about the digestive system.

In class, the teacher paired the students and instructed them to listen or view their partner's recording (Figure 4). As they listened or watched the presentation, each student would have to evaluate the parent's knowledge of the digestive system by using checklist provided to them. For example, students need to validate if the parents have indeed described the flow of food in the digestive system or demonstrated understanding of where digestion stops in the system. Learning inside and outside the classroom became a participatory activity between students, teachers and parents. Students took on the role of teaching their parents what they learnt and students learnt from their friends through their feedback on the recording. We noticed that students did exhibit accountability for what their parents had shared and they critically reflected on what they had taught or omitted during the child-parent sharing. One student in the class was concerned that he had omitted "teaching" his father a function of the digestive system. This happened after his partner said that he did not hear any mention of it from the video recording of the student's father.

#### Planned learning in class – Emergent learning out of class

In our planned lesson, we encouraged the students to inquire what they would like to know pertaining to their selected topic. For example, after students listed all that they know about animals, we asked them to list the questions they would like to know about animals. There are no restrictions on the type of questions they can pose. Using the Smartphones, they were encouraged to use the Internet to search for the answers to the questions. As young learners of inquiry, they had initial difficulties in framing the questions. They also needed support from the teacher and researchers in helping them to make sense and filter the influx of information that emerged from their search results.

In class, we promoted inquiry by asking students to list questions they have about a topic and look for evidences that would help them answer the questions. They would use the 3G-enabled phones to search for answers and share their answers in class discussions facilitated by the teacher. The planned lesson on the inculcation of inquiry skills actually spills over to learning practices at home. In one of the interviews with Roy (one of our six targeted participants for ethnographic study) and his family, his father commented that Roy developed an interest in asking questions about things around him. His mother recounted that Roy taught her that guppies give birth to young instead of laying eggs. In one of the lessons, students used the Smartphone to inquire "which fish give birth to young live?" and "which mammal lays eggs?" In class, we encouraged the development of cognitive processing skills which includes observation, recording observation on the Smartphone with the camera, and sharing with others. At home, Roy observed that the fishes in his home aquarium had laid eggs (Figure 3). He used his Smartphone camera to take pictures of the eggs. His interest and habits of inquiry was supported by his parents. On a holiday trip, Roy recorded videos of the fishes and marine life from the transparent base of the boat he was in (Figure 4). The Smartphone is a tool to support Roy's emergent learning arising from his interest in marine animals. He used the Smartphone to record his observations outside the class and search for more information. He has learnt to be observant about the world around him and started to find out answers to his questions independently. Roy told us that he shared his pictures and videos with classmates who shared similar interests in marine animals.


Figure 3. Roy taking pictures of eggs laid by the fish in his home aquarium


Figure 4. Roy taking pictures of marine life he observed from the transparent bottom of a boat

## Emergent learning in classroom - Planned learning out of classes - Planned learning in classroom

Jeremy was one of the participants in our study who showed a keen interest in exploring the use of the mobile devices when it was first introduced in the classroom. During class, he was observed to explore both the hardware and software features of the Smartphone (Figure 5). In one lesson, he used Google Maps to find the location of the school and his home on the map. His exploration was unrelated to what was taught in the class at the time. He was zooming and panning the map to locate, and making spatial comparison to the location-in-question with respect to his neighbourhood. He used Google Maps to find directions from the school to his home by entering the start point and end point of his journey. In a home interview, he shared with the researchers his motivation of using Google Maps - his grandmother had problems travelling from one place to another as she does not know the directions. He planned to use Google Maps to help her get around. Jeremy's exploration extended to the social realm as he shared with his classmates about the use of Google Maps. During class, Jeremy and another classmate used Google Maps to compare the distances of their homes from school.

Some of our ideas for learning design are inspired by the students' innovative use of mobile devices in the emergent learning spaces. For example, when the mobile devices were first introduced to the class, some students used Google Maps to find the location of their homes on the map (Figure 5) and compared the distances of their homes from school. To connect the students' emergent use of the Google Maps on the Smartphone to in-school instruction, the teacher designed a math learning activity for the students to learn distances and length in an authentic context.: Each student was asked to find the distances and directions from one point to another point (e.g. from the school to the swimming complex) by using Google Maps. They then added up the length of the journey and converted the distance from Kilometers to meters. The use of Google Maps enabled the students to have a better sense of distance as the interface displays the directions to places they are familiar with such as the swimming complex or the nearby shopping mall from the school. It enabled them to relate distance and time by viewing the estimated time needed to reach a destination using different modes of travel such as walking or driving using Google Map.

Figure 5. Jeremy exploring Google Map in class

Figure 6. Aaron viewingcollection of fishing rods

## Emergent learning outside class - Emergent learning in class - Planned learning in class

Aaron had a strong interest in fishing. His interest was fuelled by his family activities which include weekend overseas trips for fishing trips on a boat. His house was filled with equipments and artifacts which reflected the family's interest in fishing (Figure 6). During interviews, Aaron showed his fishing equipment to the researchers. At home, he used the Smartphone to view YouTube videos of the marine life. Using search terms on YouTube, he found videos that meet his interest on demand. In the classroom, Aaron shared his videos and knowledge with classmates. One of the students with whom he shared his interest is Roy. In our interview, Roy told us about Aaron telling him how fast a swordfish could swim in the deep sea. This sparked off an interest for Roy to watch "Blue Planet" and borrow books from the public library on the deep sea creatures.

Aaron's interest in the marine life extended to his knowledge about animals. In a lesson where the teacher asked students to animate the characteristics of the mammals, Aaron drew an Elephant Shrew. He correctly identified the characteristics of a mammal. His choice of drawing an Elephant Shrew showed his knowledge of animals. He asked questions about fishes such as "Where do Red Groupers live?", "Where do Water Hares live?" and "How do fish lay eggs?" in his KWL chart (What do I know, what do I want to know, what have I learned) in the Smartphone. His knowledge about fishes is impressive for an 8 year old as he stated that "Rabbit fish live in holes in the water", "fish have gills to breathe in water", and "fish use their fins to swim." In class, he continued to increase his understanding and interest using the Smartphone to look for information and view videos.

## Students' Attitudes towards Smartphones for Learning

The scenarios described above provided us an understanding of how individual student learn in different contexts, in and out of the classroom, and across different social contexts over time. This project seeks to examine changes in the 39 students' attitudes towards learning with the Smartphone in and out of classroom after introduction of the devices and designed lessons. In the survey, the students provided feedback on the affordances of the Smartphone and its role to helping them to learn after six months of usage. The survey results are presented next.

Among the 39 students, around 80% of the students feel Smartphone helps their learning. 85% agree that the Smartphone help them learn class subjects and 79% agree that Smartphone help them learn things outside of school. 62% students agree that they understand better the science concepts learned in class and how things they learn in class are connected to their daily lives.

Table 2. Students' attitudes towards HTC phone & its role for learning (N=39)

| | Agree (%) | Neutral/ Not sure (%) | Disagree (%) |
|---|---|---|---|
| It is easy to use. | 71.8 | 20.5 | 7.7 |
| It is light enough for me to carry. | 74.4 | 17.9 | 7.7 |
| The size of the screen on the HTC Smartphone is too small to do my school work. | 25.6 | 10.3 | 64.1 |
| The size of the keyboard on the HTC Smartphone is too small to do my school work. | 17.9 | 15.4 | 66.7 |

| | | | |
|---|---|---|---|
| I do not think that using the Smartphone helps my learning | 2.6 | 17.9 | 79.5 |
| It helps me learn my class subjects. | 84.6 | 12.8 | 2.6 |
| It helps me learn things outside of school. | 78.9 | 18.4 | 2.6 |
| It distracts me from doing my school work. | 7.7 | 25.6 | 66.7 |
| I understand the science concepts learned in class better | 61.5 | 35.9 | 2.6 |
| I understand better how things I learn in class are connected to my daily life. | 61.5 | 33.3 | 5.1 |

When comparing students' attitudes toward the role of Smartphone in their learning in September survey with their attitudes in the February survey, we found that from the paired-sample *t* test, students' attitudes have a positive change towards the use of mobile devices for learning in class ($t(38)=-2.765$, $p<.01$) and out of class ($t(38)=-2.321$, $p<.05$) after the introduction of the smartphone and a series of "mobilized" lessons (Table 3). Students like the learning activities using computers and gadgets more than before ($t(38)=-2.016$, $p<.05$).

Table 3. Results of paired-sample t test on students attitude (N=39)

| | | Mean | Std. Deviation | *t* |
|---|---|---|---|---|
| Mobile device helps me learn my class subjects. | Feb Survey<br>Sep Survey | 1.46<br>1.82 | .643<br>.451 | -2.765** |
| Mobile device helps me learn things outside of school. | Feb Survey<br>Sep Survey | 1.42<br>1.76 | .683<br>.490 | -2.321* |
| I like the learning activities using computers and gadgets. | Feb Survey<br>Sep Survey | 1.05<br>1.23 | .223<br>.536 | -2.016* |
| I learn more when I work in a group than alone. | Feb Survey<br>Sep Survey | 1.37<br>1.68 | .633<br>.662 | -2.634* |

*$p<.05$, **$p<.01$.

Both survey questions asked students if they think they can learn by "playing video and computer games", "watching videos on the computer or the web", and "browsing the Internet". Before the HTC phone introduction, only 26% students feel they can learn by "playing video and computer games", 38% of them think they can learn by "watching videos on the computer or the web", and 65% think they can learn by "browsing the Internet". Half a year later after the introduction of HTC phone, the percentage increase to 56%, 87% and 92% respectively. After 6 months of using the Smartphone, a student gave the following feedback in the survey:

> I like to have more projects on My Projects and have more Sketchy activities. Sketchy helps me learn to raw (draw) nicer, write nicer, it helps me to search for more information from the Internet….The Internet helps me to search for pictures I like and searching for more information.

The use of the Smartphone facilitates the students learning and fulfills their own personal interest. Another student commented: "I like using because it has games, worksheets and teaches us more about animals, fungi, materials, length and many other more." A student indicated her interest in using the Smartphone outside of school when she said "I prefer to use the phone outdoors like in the zoo or park."

## Conclusion

This paper discusses how the affordances of mobile technology match the premises of seamless learning, and present seamless learning scenarios from our research. Till recently, not many studies in the literature make full use of the vast potential for learning spaces out of classroom to enrich school education. Studying in and out of class learning will enable us to gain insights and build theory on the notion of seamless learning. As discussed in this paper, students' learning spaces can be enhanced or extended by mobile technologies which support learning across multiple contexts and time scales. Mobile technology can support seamless learning in a variety of contexts over long periods of time. By enabling learners to learn 'anytime, anywhere', mobile technology augments the propensity for students to engage in self-directed learning on the fly. Student could adapt the use of mobile technology in different contexts for different purposes that could be related to their interests or learning goals. The mobile technology enables students to learn, through ways of organizing ideas, memories and personal resources, to support seamless learning. It enables students to describe and preserve their observations and reflections, by means of diaries, planners, timelines, picture taking and audio/video recording etc.

From our initial findings, we observed how learning occurs and is connected seamlessly across the boundaries through social interactions and the environment mediated by the use of the technology. Students view the Smartphones as learning tools as they use it over time inside and outside the classroom. Understanding technology-supported seamless learning is important for researchers and practitioners who are interested in connecting classroom learning and out of classroom learning spaces in order to create rich and holistic learning experiences for the students. We hope this paper will stimulate further discussions on how to design and study seamless learning environments that can foster 21st century knowledge and skills among the young generation of learners.

## References

Barron, B. (2006). Interest and self-sustained learning as catalysts of development: a learning ecologies perspective. *Human Development*, *49*, 4, 193–224.

Chan, T.-W., Roschelle, J., Hsi, S., Kinshuk, Sharples M., *et al* (2006). One-to-one technology enhanced learning: an opportunity for global research collaboration. *Research and Practice of Technology Enhanced Learning*, *1*, 3–29.

Chen, W., Seow, P., So, H.J., Toh, Y, & Looi, C.K. (2010). Connecting learning spaces using mobile technology. To appear in *Education Technology*, September-October, 2010.

Goldman, R. (2004). Video perspectivity meets wild and crazy teens: a design ethnography. *Cambridge Journal of Education*, *34*(2), 157-178.

Looi, C.K., Seow, P., Zhang, B.H., So, H.J., Chen W., & Wong, L.H. (2010). Leveraging mobile technology for sustainable seamless learning: A research agenda. *British Journal of Educational Technology, 41*(2), 154-169.

Looi, C.K., Wong, L.H., So, H., Seow, P., Toh, Y., Norris, C., Soloway, E., Chen, W., & Zhang, B.H. (2009). Anatomy of a mobilized lesson: Learning My Way. *Journal of Computers & Education, 53*(4), 1120-1132.

Pea, R. (2009). Fostering learning in the networked world: Trends, opportunities and challenges for learning environments and education, *Keynote speech presented at 3rd Redesigning Pedagogy International Conference*. Singapore: National Institute of Education.

Roschelle, J. (2003). Keynote paper: unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning*, *19*, 3, 260–272.

Sharples, M., Taylor, J., & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews & C. Haythornthwaite (Eds.), *The Sage Handbook of E-learning Research* (pp. 221-247). London: Sage.

So, H. J., Kim, I. S., & Looi, C. K. (2008). Seamless mobile learning: Possibilities and challenges arising from the Singapore experience. *Educational Technology International, 9*(2), 97-121.

Squire, K. & Klopfer, E. (2007). Augmented reality simulations on handheld computers. *Journal of the Learning Sciences*, *16*(3), 371–413.

Stroup, W. M. & Petrosino, A. J. (2003). An analysis of horizontal and vertical device design for school-related teaching and learning. *Education, Communication & Information*. Vol.3 (3). pp.327-345.

Zhang, B. H., Wong, L. H., Seow, P., Chen, W., & Looi, C. K. (2009). "Mobilizing" curriculum materials for seamless learning—Opportunities and challenges. Paper presented at the *Global Chinese Conference on Computers in Education,* Taipei, Taiwan.

## Acknowledgements

This material is based on the work supported by the National Research Foundation (Singapore) under Grant NRF2007-IDM005-MOE-008. We also wish to thank Nan Chiau Primary School for collaborating with us on this research, and Baohui Zhang, Gean Chia and Ivica Boticki for working on various aspects of this project.

# A Tempest in a Teapot Is but a Drop in the Ocean: Action–Objects in Analogical Mathematical Reasoning

Dor Abrahamson
University of California—Berkeley, Berkeley, CA 94720–1670
dor@berkeley.edu

**Abstract:** We discuss a brief transcribed excerpt from a task-based interview with Li, an 11.5-year-old participant in a design-based research study of probabilistic cognition pertaining to the binomial. We investigate whether and how Li made sense of the behavior of an unfamiliar computer-based artifact—the diminishing proportional impact of successive random samples on the overall shape of a dynamically accumulating outcome distribution. Li constructed two informal analogical situations as multimodal discursive means of concretizing, elaborating, and communicating his emerging understanding of the artifact's behavior. These non-routine utterances shifted the discourse to an explicitly embodied, imagistic space bearing unique affordances for negotiated epistemic syntheses of phenomenological and technological constructions of quantitative relations. Microgenetic analysis suggests that Li's presymbolic notion was not a static magnitude but an intensive-quantity "action–object"; he subsequently unpacked this dynamical "*a*/*b*" qualia into its constitutive "*a*" and "*b*" elements. We reflect on implications of this counter-curricular sequence for educational design.

## Background: Student Analogy as Researcher Opportunity

We are interested in the phenomenon of mathematics learning. We conceptualize mathematics learning as the process in which an individual builds meaning for mathematical artifacts, such as notions, semiotic devices, and procedures. We research this learning process in an attempt to understand what it is that students do when they build meaning for mathematical artifacts, what teachers do to support this process, and what instructional materials may best serve this process. In the current study, we examine one student's metaphorical reasoning as he attempts to construct meaning for the behavior of an unfamiliar mathematical artifact. In particular, we seek to determine the student's initial ontology of the artifact as indicating his pre-articulated phenomenological resources that could plausibly serve as proto-mathematical. So doing, we question the warrant of some epistemological assumptions underlying traditional mathematics curriculum.

Our comments in this essentially theoretical paper should be taken as no more than conjectural. Nevertheless, an appeal of the paper would be the potentially effective theoretical fit, and hence methodological fit, between the dynamical nature of the particular artifact at the center of the student's inquiry and our interest in the dynamics of multimodal mathematical reasoning. This fit enables us to elaborate on Phenomenology tenets and Learning-Sciences conjectures regarding the precedence of unreflective perception of action over analytic construction of concepts, as we investigate the purchase of these conjectures on our empirical data. Ultimately, we hope to draw tentative conclusions aligned with parallel research efforts so as to offer an intellectual space for conversation informing future studies.

Probability—the mathematical content selected for this study—is amenable for our investigation of tension between phenomenological and cultural constructions of situations and artifacts. That is, the unique epistemological "mode" created by situations involving uncertainty, and in particular the challenges of articulating such uncertainty symbolically, appears to impel students to seek non-symbolical discursive genres as means of expressing their intuitive quantitative reasoning (e.g., Rubin & Hammerman, 2007). The consequent protracted discursive interim from the embodied to the symbolical is conducive for examining the microgenesis of meaning. Namely, we conjecture that the metaphors generated by this study's focal student served him as more than vehicles for communicating "ready-made" coherent notions. Rather, we believe that the cognitive–discursive actions of evoking and "grammarizing" these metaphors were instrumental for the student's *initial* articulation of his reasoning. If this conjecture bears out through further research, it would present a number of implications for mathematics-education theory and practice. In terms of theory, we would suggest that idiosyncratic metaphors could enable diverse individual students to ground challenging mathematical constructs in their prior quantitative experiences, both formal and informal. In terms of practice, we would encourage teachers to cultivate socio-mathematical classroom norms of discourse in which such metaphorical constructions are approved and possibly even solicited.

## Metaphor as a Unique Semiotic Means of Objectification

When students encounter a new situation and attempt to determine what it is, they are tacitly attempting to determine what it is *like*. This reasoning process is usually opaque to others as well as, perhaps, to the students. Usually, it is only when the students subsequently respond in non-normative ways to problem situations that we assume that their underlying conceptualizations of the situation are non-normative. By stating their metaphors explicitly, though, students create opportunities to reflect more concretely on their emerging conceptual system for a given mathematical subject matter content as well as to receive targeted formative assessment and guidance in learning this content. We conjecture that instructional discourse around metaphor plays a unique role in students' bridging between tacit and cultural constructions of quantitative situations. Through analyzing a student's metaphorical constructions, we hope to promote research into this conjecture. Our proposal to endorse metaphorical reasoning as central to the practice of mathematicians and, hence, of mathematics-education researchers, agrees with recent emerging interest in "embodied" or "multimodal" forms of quantitative reasoning (Abrahamson, 2009b; Edwards, Radford, & Arzarello, 2009; Lemke, 1998; Presmeg, 2006) as well as with converging testimonies and theory pertaining to the explorative and desultory process of mathematical learning and discovery (Lakatos, 1976; Schoenfeld, 1991; Thompson, 1993). Furthermore, demonstrating pivotal roles of multimodal mathematical reasoning would add to the ongoing conversation among cognitive scientists regarding a looming paradigm shift from amodal to modal conceptualization of human reasoning (see Barsalou, 2008; but cf. Dove, 2009).

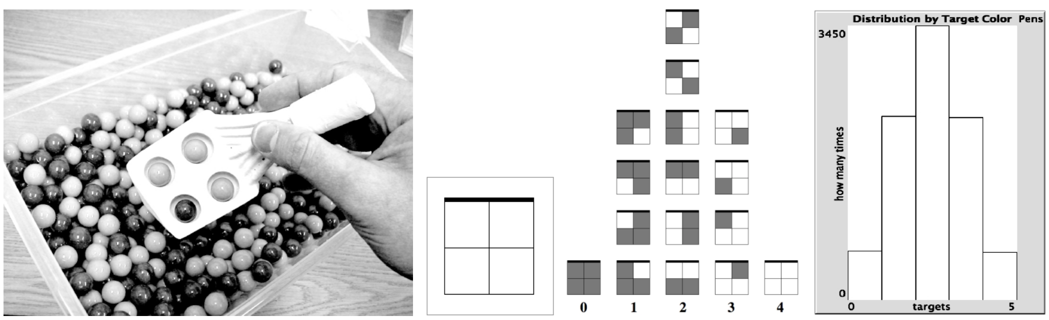
We view mathematical learning as a dialogic, distributed, reflexive, and emergent process. Students learn by building personal meanings for mathematical artifacts. They do so by appropriating these artifacts as significant means for multimodal semiotic objectification of their presymbolic notions. It is thus that designed and guided participation in artifact-based social activities mediates disciplinary knowledge as internalized discursive praxis (Mariotti, 2009; Radford, 2008; Sfard, 2002). And yet, students' naïve presymbolic notions, judgments, implicit heuristics, and inferences for quantitative situations are sometimes more sophisticated than they can initially express (e.g., Gelman & Williams, 1998), particularly when students who are not yet fluent in proportional constructs attempt to express notions of perceptually privileged intensive quantities (Piaget, 1952; Stroup, 2002). Thus, it is important that educators create opportunities for students to avail of their capacity for pre-articulated quantitative reasoning, such as by making available in the learning environment mathematical artifacts that accord with structure information students tacitly seek (Abrahamson, 2009b). In this negotiation of meanings for mathematical artifacts (Cobb & Bauersfeld, 1995), metaphor serves as a powerful discursive engagement, because it explicitizes—for the student, teacher, classroom, *and researcher*—the student's (idiosyncratic) reasoning.

That said, for this study we do not take a philosophical or cognitive position as to how students first develop these metaphors or how they evoke them in situ as inferential discursive mechanisms (cf. diSessa, 1983; Halliday, 1993; Lakoff & Núñez, 2000; see in Ortony, 1993). Instead, we treat the evocation of metaphor at a larger analytic granularity, as a species of abductive inference amid encounter with a novel problem (e.g., Prawat, 1999). Moreover, we are committed to framing students' reasoning as situated not in a solipsistic void but as an integral aspect of discourse (Sfard, 2002). Particularly, we are inspired by Radford's (e.g., 2008) view of mathematics learning as guided discursive processes in which students appropriate semiotic artifacts as means for objectifying their presymbolic notions. From this semiotic–cultural perspective, we wish to examine the hypothesis that metaphor, too, is a means of objectification and to explore its unique properties, roles, and prospects within institutionalized practices of mathematical learning. We propose that students recall personal experiential gestalts as a means of reifying the emerging coherence they sense in mathematical artifacts yet initially lack formal constructs or vocabulary to express.

## Source Data and Research Questions

The transcribed excerpt that constitutes the source data for our study is from an interview with Li, an 11.5 year-old male student ranked by his mathematics teachers as high-achieving. Li was one of 28 Grades 4–6 study participants from a private suburban K–8 school in the greater San Francisco East Bay area. The context of these 1-hr. interviews was a design-based research study of late-elementary and middle-school students' intuitive notions pertaining to the mathematical study of probability as well as the potential roles that a set of mixed-media technologies of our design could play in enabling these students to leverage any such naïve notions toward understanding mathematical formulations of random phenomena, specifically the binomial. Using a flexible protocol, we conducted semi-structured clinical interviews (Ginsburg, 1997). The sessions were videotaped for subsequent analysis, and selected episodes were transcribed. The research team employed collaborative qualitative microgenetic analysis techniques (Siegler & Crowley, 1991) as well as grounded theory (Glaser & Strauss, 1967), through which emergent insights were articulated and iteratively crosschecked against the entire data corpus as a means of consolidating, investigating, and developing new constructs (diSessa & Cobb, 2004). Instances of student elaborate metaphorical reasoning were extremely rare (only four identified within a total of 26 hours of data) yet appear to be uniquely informative. In particular, students' gestures therein enable us to speculate on their conceptualizations of proportionality and probability (e.g., Alibali, Bassok, Olseth, Syc, & Goldin-Meadow, 1999).

Figure 1 overviews four central objects and target artifacts that were used, constructed, and/or automatically generated during the interview. The marbles scooper is used to draw a sample of exactly four ordered marbles from an "urn" of mixed green and blue marbles (see Figure 1a). This hypergeometric experiment—*during* each 4-marbles sample draw, marbles are not replaced—approximates the binomial due to the ratio of *n* (4 marbles) to the content of the urn (hundreds of marbles). In the interview, we first ask the participant to guess what would be the experimental results of scooping. Next, the participant is guided to use crayons and a pile of stock-paper cards each bearing a blank 2-by-2 matrix (Figure 1b), so as to create the sample space of the experiment and assemble it in the form of the *combinations tower* (see Figure 1c). Later, the dyad engages computer-based simulations of the same experiment (e.g., see interface fragment—a histogram—in Figure 1d). Here, we examine one student's reaction to one feature of one artifact—Li's reaction to the diminishing proportional impact of samples in the "4-Blocks" simulation.(1)




*a.* *b.* *c.* *d.*

Figure 1. Materials used in the study—theoretical and empirical embodiments of the 2-by-2 mathematical object: (a) The marbles scooper; (b) a template for performing combinatorial analysis; (c) the combinations tower—a distributed sample space of the marbles-scooping experiment; and (d) an actual experimental outcome distribution produced by "4-Blocks," a computer-based simulation of this probability experiment.

The computer-based model "4-Blocks," built in NetLogo (Wilensky, 1999), simulates the marbles-box probability experiment. 4-Blocks includes a virtual 2-by-2 array (see Figure 2, next page). When a virtual sample is taken, each of the four cells in this array is randomly assigned either green or blue coloration. For example, the sample may be three green and one blue cells in any of the four possible orders (hence, "3g1b"). The model computes the number of green cells in the array, such as 3, and this value is supplemented to a list (that by default is not seen by the user). Immediately, the interface reflects this new result by "bumping up" the appropriate histogram column, such as the second-from-right column, by a vertical extent commensurate with one unit. A feature of 4-Blocks is that its histogram updates dynamically even as the experiment is running. (To interact with the 4-Blocks model as well as to view the video clip discussed herein, visit http://edrl.berkeley.edu/publications/conferences/ICLS/Abrahamson-ICLS2010/.)

The vertical extent of a unit is dynamically calibrated to the histogram's maximum *y*-axis value (e.g., this *y*-value is 43, in Figure 2). That is, the histogram has an "autoplot" feature: when a category column "grows" such that it is about to exceed the maximum *y*-value, this value increases by one unit so as to accommodate the impending growth. The column thus remains as tall as it was, "glued" to the top of the histogram frame, while other columns appear to sag down a unit. Consequently, as the simulation is running, the vertical extent of a single unit keeps diminishing. Thus, when the maximum *y*-value is 10 (the default initial value), a unit is 1/10 of the height of the frame, resulting in the perception of a major upward jolt on the screen when a sample is taken. In Figure 2, a unit would be smaller, at 1/43 of the height. When the maximum *y*-value is relatively large, such as 1,000, a unit is only 1/1,000 of the histogram, resulting in a minute perceptual change. Moreover, at 10,000, an upward motion may demand finer calibration than the pixels could accommodate, so that very often the columns would not register any perceivable change at all.

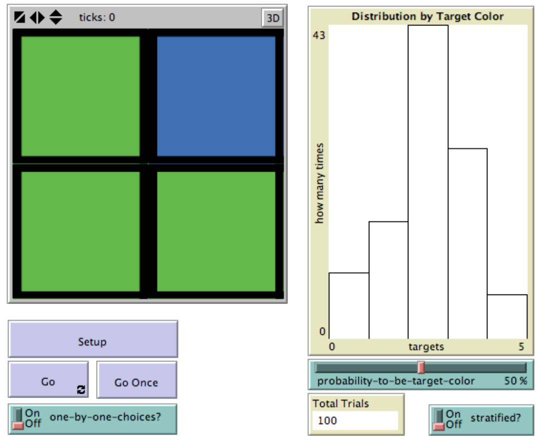



Figure 2. The 4-Blocks NetLogo simulation of the marbles-box probability experiment after 100 samples.

This phenomenon—the diminishing perceptual impact of samples as the experiment runs—is a potentially useful feature of the NetLogo modeling-and-simulation environment, because of its inherent capacity to gradually shift the user's attention away from the impact of each haphazard sample and toward the distribution gestalt, which the user might thus objectify as an aggregate property of the simulated phenomenon (Wilensky, 1997). Moreover, the absence of numerals along the histogram's *x*- and *y*-axes appears to impact the user's primary construction of relations among the histogram columns: rather than constructing these relations as mediated through symbolically notated absolute values and calculated differences among them, the user can construct them proportionately based on immediate perceptions of relative heights and/or areas. Thus, in each and every long-term run of the simulation, the user may experience opportunities to witness and reflect on the gradual yet ineluctable emergence of the distribution shape, such as the 1-4-6-4-1 shape of the marbles-scooping binomial experiment set at $p = .5$ (Figure 1d); with guidance, users may be able to construe this shape as objectifying their presymbolic sense of the phenomenon's aggregate characteristics (e.g., the expected plurality of 2g2b, see Abrahamson, 2009b).

By and large, such was the case with all our 28 participants, including Li. Yet how would Li make sense of the histogram? In particular, would Li perceive the impact of each sample upon the distribution as proportional to the quantity of aggregated outcomes? If so, what resources might this student, who has not studied ratio-and-proportion formally, bring to bear in making sense of this phenomenon of diminishing perceptual impact, and how might he articulate his inferences? From an instructional-design perspective, is the technologically sophisticated semiotic system built into the autoplotting histogram helping or hindering Li's learning?; is this histogram "cognitively ergonomic" (see in Abrahamson, 2009b), and if so, which phenomenological resources could it tacitly cue and how might Li articulate these notions mathematically?

## Analysis: Analogy-Based Unpacking of Intensive-Quantity Action–Objects

In this section we present and discuss a 41-second videographed excerpt and annotated transcription of a conversation fragment between Li and the interviewer, in which Li reasons analogically about the samples' diminishing perceptual impact on the histogram shape. The excerpt begins when Li and Dor have already negotiated the construction of the sample space, have assembled it in the form of the combinations tower, have discussed relations between the marbles-box experiment and the combinations tower, and have been working on 4-Blocks. Doing so, Li draws on non-mathematical phenomena as contexts for his quantitative reasoning. Li's rhetorical strategy is to build an argument by comparing two extreme cases of sampling: (a) at the beginning of the experiment, when a relatively small number of outcomes has accumulated so that each supplementary sample causes quite a "splash"; and (b) well into the experiment, when a considerable number of outcomes has accumulated, so that any additional sample causes but a "ripple." His first context—the glass/lake analogy—will present this contrast in full, and then the second context—batting averages—elides the first of the two cases in pragmatic enthymeme and states only the second case. Just prior to these analogies, though, Li anticipates that the emergent experimental frequency distribution will resemble the combinations tower in shape; he will then monitor this emergence to evaluate his prediction.

Dor: So what do you think might be the shape of the **columns**, as **it** goes up?
Li: [Gesturing to the combinations tower] Something like this.
Dor: Well let's see. [activates experiment, explains autoplotting] So what's happening?

Li: [Gesturing to the distribution] **It's hovering at around this** [gestures to combinations tower]…. [20 seconds later, when 5,000 samples have been drawn] **Look, it's** [the frequency distribution] **almost exactly like this** [the combinations tower]…. [30 seconds later] **Now they're** [columns] **moving less** [inaudible].

Dor: Why is **it** moving less?

Embedded in the interviewer's questions is an implicit linguistic passage from "the columns" to "it." Li picks up this cue and accordingly refers to the columnar frequency distribution with the pronoun "it." Irrespective of the fullness of his understanding at this point, Li's appropriation of the interviewer's singular pronoun as reference to the column collective suggests he is construing the distribution as an intact, if amorphous, object. Soon after, Li again frames the aggregate motion as "plural" behavior. Below, this *motion* is about to become the *object* of discussion. Namely, the histogram's figural change is first described with a verb ("moving") but will soon become a noun ("splash" vs. "ripple") (cf. Bakker, 2007, for a similar case of Peircean 'hypostatic abstraction' in statistical analysis). What is unique about these nouns is that they are inherently about change. That is, the object at the core of Li's proto-mathematical reasoning is not a static magnitude, such as the measured vertical displacement of the histogram bars (an '*a*' element) or the cumulative number of samples (a '*b*' element)—Li's tacit phenomenological primitive in this dynamical experience is a synthetic a priori—an intact action-based *a*/*b* change-over-time intensive-quantity unit (cf. Stroup, 2002). Thus, technological affordances of visualization media—the dynamical autoplotting histogram—may reverse mathematics learning sequences from the traditional magnitudes-before-relation to relation-before-magnitudes. Li's ensuing metaphorical outburst, in which he deftly unpacks the intact *a*/*b* change–object into its *a* and *b* constituents, we submit, implies that we should take pause, as mathematics educators, to consider the possibility that students learning intensive quantities, such as in the mathematics of change, may avail of trajectories that go counter to traditional curricular sequencing. It is thus that computers may "restructurate" mathematical content (Wilensky & Papert, 2009).

Yet what phenomenological resources will Li bring to bear in explicating the change–object? Whereas the NetLogo histogram, when the model is run, foregrounds change as the salient object of attention and may thus render the mathematics of change more accessible to learners, the autoplotting feature may conceivably challenge students due to its ostensible phenomenological aberration. That is, in "real life," we may muse, growing aggregates actually occupy greater space. Intriguingly, however, natural visual perception is a perspectival experience, in which retinal prints of objects depend on their distance:

Li: /2 sec/ **Because**…/3 sec/ [gazes up to the wall] **the larger number**…/3 sec/ uhh… ["checks in" with the interviewer] /2 sec/ [rapidly] **Like if you have a little glass** [iconic gesture: LH cups a glass in natural position near body; gazes at glass] **of water and you drop a marble in** [iconic: LH uncups, rises, drops marble], **it's gonna be... there's gonna be, like, a splash** [LH, palm up, abrupt vertical rise], **but if you have a big giant lake** [LH, palm up, drawn back and up above head to encircle expansive lake; RH, in jacket pocket, budges to complete circumference], **and you throw a marble in** [LH catapults marble, then scratches nose]**, there's just gonna be a ripple** [joins LH thumb and index, lowers hand to chest height on right-side of embodied space; taps fingers together, possibly marking marble's contact with surface, then inscribes smooth horizontal line across to the left; hand opens]. **It's a, it**… [gestures to histogram, orients gaze and pivots body towards it]

Dor: **Oh, ok. Like each individual additional sample** [LH gyrates swiftly, iterating addenda] **is causing less of a** [LH & RH "contain" combinations-tower outline]**…**

Li: [cuts in mid sentence] **Yeah, it's like a batting average in baseball. If you've already had five hundred at-bats**(2) [LH opens, shifts slightly to the left, palm up, "holding" the 500 at-bats, then relaxes]**, and then you get out one more time, it's not going to make it go down that much—it'll make it go down like two points** [LH pinches thumb and index, as per "ripple," as if to subtract two points]**, or**…

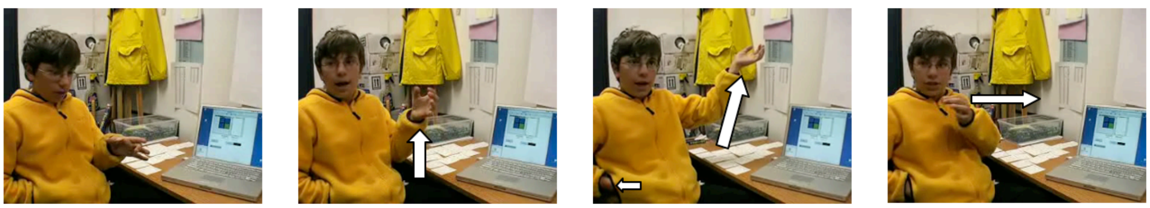
Dor: Ok, so as we go along, each successive sample causes less of a commotion.

Li: **Yeah.** [Gazes back at the computer screen] **Look, they're just barely moving.**

The conversation, above, pertained to variation in the effect of a single sample on the proportional distribution of experimental outcomes: as outcomes aggregate, the proportional impact of each additional sample diminishes. Viewed on the computer screen, the self-compressing histogram converges on the 1-4-

6-4-1 distribution, eventually portraying the dynamic aggregation as perceptually static, albeit the constantly updating maximal *y*-axis value reveals the process as additively active. Li appears to understand this principle. However, this high-achieving $6^{th}$-grader's limited fluency with proportional constructs does not enable him to capture the process with appropriate vocabulary, such as "proportion" and its cognates, or with suitable mathematical constructs and arithmetic operations for conceptualizing and treating rational numbers, as witnessed in his aborted attempt, "Because the larger number…" Consequently, Li evokes a situation analogous to the dynamic artifact yet evaluated as more conducive to objectifying the principle. Li thus unpacks "splash > ripple" as "one-marble-impact : size-of-glass > one-marble-impact : size-of-lake."

The key physical dimension underlying the splash-vs.-ripple articulated comparison is the vertical height of the water displacement in each. And yet, would not a marble thrown into a lake produce a vertical displacement of water that is at least as large if not larger than in a glass? What is at stake here, it appears, is not the absolute physical size of phenomena but their pre-rationalized perturbations to the perceptual field. Thus, ratio is pre-built into perspective as an optical calibration. Moreover, equally sized distal events are normalized by their containers, because these perceptions' phenomenological circumstances are such that events occurring in larger containers are physically farther from the viewer and therefore retinally smaller. Li's spontaneous gestures inscribe the glass and lake *as they would be perceived in the unreflective phenomenology of lived experience*, so that the lake occupies little more optical canvas than the glass does upon Li's "visuo–spatial sketchpad" (see Baddeley & Hitch, 1974).[(3)] Thus, Li can compare two intensive quantities—the splash and the ripple—on the basis of their retinal magnitude alone (see Figure 3, below).[(4)]



Glass Splash Lake Ripple

Figure 3. Dominant gestures in Li's glass-vs.-lake comparison analogy, with arrows as action overlays.

In summary of Li's first analogy, his argumentation responded to the interviewer's request for an explanation regarding the patterned behavior of a dynamical mathematical object—an electronic histogram that is confined in its total interface real-estate yet monitors an ever-aggregating vertically surging outcome distribution. Li's analogical constructions, however, *shifted the dyad into an imagistic discursive space, wherein larger aggregates are in fact embodied in larger objects*. Therein, Li compared two intact experiential events, both a priori intensive quantities; only through engaging in thinking-for-speaking (Slobin, 1996) did Li unpack each of these quantities "cubistically" (Nemirovsky & Ferrara, 2009) such that its respective constituents were "liberated" (Bamberger & Ziporyn, 1991) into *a* and *b* magnitudes so that each event was reconstructed as an *a*/*b* qualitative quotient. This unpacking is anecdotally enhanced by Li's essentially unidextrous sequential gesturing, which radicalizes the discursive caveat of linear utterance.

In selecting a baseball context for his second analogy, Li evokes a situation that is more conducive to mathematizing his argument. First, monitoring quantitative aspects of this popular activity is a familiar, culturally appropriate practice. Second, the baseball example is more ontologically consistent than the water example: in the glass-vs.-lake analogy, the perturbation was caused by a marble dropped into water and thereafter the marble is ignored, whereas in the baseball scenario, the perturbation was caused by the player performing an at-bat, which is thereafter added to the denominator of the "batting average" (and to the numerator, contingent on how the at-bat played out). Third, the baseball analogy offers a pre-enumerated set of discrete units, 500 at bats, whereas the water analogy would demand measurement of the continuous body before further calculation could be performed. Finally, a batting average is an a priori proportional construct, so that this analogy is much nearer to the probability context—culturally, semantically, semiotically, and arithmetically—than the water-surface analogy. Thus, in selecting the baseball analogy, Li also steers the dialogue from the rich imagistic detour, which the interviewer had countenanced, back to the institutionally normative genre of discussing quantitative reasoning numerically. Using numbers as precise signifiers of quantity, in the baseball analogy, Li need not communicate magnitudes gesturally—indeed, comparison of the vast "big giant lake" gesture to the modest "500 at-bats" cupping as well as, analogically, comparison of the lake-impact gesture to the "-2" (read: "-.002") gesture demonstrate pragmatic contraction in iconicity from the first to the second context (see Radford, 2008).

Having completed his two-analogy argumentation sequence (see Figure 4, below), Li looped the dyad back to the phenomenon under inquiry. Presumably, the interlocutors, who renew their co-attention to the histogram, do so with newly shared professional vision established in analogical discursive space.[5]

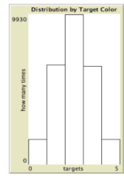
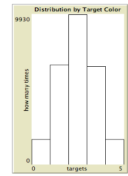
| **Impact** | MAJOR | | | BARELY DISCERNABLE | | |
|---|---|---|---|---|---|---|
| **Context** | **Before** | **Perturb** | **Effect** | **Before** | **Perturb** | **Effect** |
| *Marbles-Scooping Experiment* |  | +1 sample → |  |  | +1 sample → |  |
| | ↕ | | ↕ | ↕ | | ↕ |
| *Marble in Water* | Cup | throw marble | "Splash" | Lake | throw marble | "Ripple" |
| | | | | ↕ | | ↕ |
| *Batting Average* | – | – | – | Batting average at 500 at-bats ($x$/500) | One out | Batting average at 501 at-bats ($x$/501) |

Figure 4. Analogizing the proportional impact of an addend as a function of the size of the aggregate.

Li's imagistic analogical detour suggests that he needed to step back from the sophisticated semiotic device and reconstruct a phenomenological scaffold by which to perform epistemic adjustment that then enabled him numerical objectification of his presymbolic image of proportional convergence. Implicit to this process of mathematization was re-describing multimodal phenomenology in conventional techno–scientific form. This bridging task, an aspect of meta-representational capacity (diSessa & Sherin, 2000), is cognitively non-trivial. We implicate the technologically sophisticated self-adjusting histogram as complicit to the tension between Li's tacit and cultural presentation of quantitative information: by perceptually foregrounding proportional change over additive change, the autoplotting histogram offers a cognitively ergonomic engagement of the mathematics of change, yet to avail of this restructurated entry to the disciplinary practice, students require guided opportunities to negotiate embodied and inscribed constructions of focal phenomena. Metaphorical reasoning is one means of accomplishing this negotiation.

## Conclusions

Whereas there is a certain logical appeal to thinking of an object as somehow cognitively simpler than an action (object + motion), the phenomenological perspective perceives actions as prior to objects—objects need to be deliberately pulled out of experience as transcendental to unreflective action. Analysis of Li's imagistic construction revealed that he was drawing on phenomenological gestalts that are experientially a priori to mathematically articulated notions of ratio. Understanding such processes could inform the work of mathematics educators. In particular, educators who reason about mathematics logically but not phenomenologically are liable to introduce $a/b$ constructs by initially presenting $a$ and $b$ as separate elements and only then calculating their quotient and suggesting meanings for this number. Yet such would-be scaffolding may sometimes hamper rather than optimize students' opportunities to draw on their informal reasoning, particularly their tacit intensive-quantity gestalts, in constructing mathematical notions. Instructional design solutions, we believe, lie in between, in creating opportunities for students to negotiate phenomenological qualia and their mathematical reconstructions (Abrahamson, 2009a, 2009b, 2009c).[6]

## Endnotes

(1) Elsewhere, we furnish detailed explanation of the design motivation and rationale and report on findings and design modifications throughout the iterated study cycles (Abrahamson, 2009c). Therein, we also report on the earlier part of Li's interview, focusing on his "semiotic leap" from presymbolic to articulated notions.
(2) A "batting average" is the cumulative ratio of hits (successful batting) to at-bats (opportunities to do so).
(3) It could have been fascinating to witness how Li would then inscribe these images using pencil and paper.
(4) The elongated horizontal extent of the ripple gesture does not map onto any *mathematical* analog—it is not the $a$, the $b$, or the $a/b$—yet it is what one would see/feel, so it is gestured. Paradoxically, this mathematically excessive magnitude could constitute a disservice to the comparison, because it inflates the $a$. Yet this "soothing" gesture may be contributing a *holistic* diminutive affect that thus in fact mitigates the ripple in comparison with the splash.
(5) The interviewer never appeared perturbed by dimensions of embodiment, discourse, and inscription implicit to Li's utterance, because he shares with Li tacit cognitive mechanisms for engaging multimodal reasoning.

(6) The data analyzed herein were collected in a study supported by an NAE/Spencer Postdoctoral Fellowship 2005-6. The ideas we present herein are the result of the Seeing Chance group's collaborative work, foremost Mike Bryant

## References


Abrahamson, D. (2009a). Embodied design. *Educational Studies in Mathematics, 70*(1), 27-47.

Abrahamson, D. (2009b). Orchestrating semiotic leaps from tacit to cultural quantitative reasoning. *Cognition and Instruction, 27*(3), 175-224.

Abrahamson, D. (2009). A student's synthesis of tacit and mathematical knowledge as a researcher's lens on bridging learning theory. In M. Borovcnik & R. Kapadia (Eds.), Research and developments in probability education [Special Issue]. *International Electronic Journal of Mathematics Education, 4*(3), 195-226. Accessed Jan. 1, 2010 at http://www.iejme.com/032009/main.htm.

Alibali, M. W., et al. (1999). Illuminating mental representations through speech and gesture. *Psych Science, 10*, 327-333.

Baddeley, A. D., & Hitch, G. J. L. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). NYC: Academic Press.

Bakker, A. (2007). Diagrammatic reasoning and hypostatic abstraction in stats. ed. *Semiotica, 164*(1/4), 9-29.

Bamberger, J., & Ziporyn, E. (1991). Getting it wrong. *The World of Music, 34*(3), 22-56.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Cobb, P., & Bauersfeld, H. (Eds.). (1995). *The emergence of mathematical meaning*. Hillsdale, NJ: LEA.

diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 15-33). Hillsdale, NJ: LEA.

diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sciences, 13*(1), 77-103.

diSessa, A. A., & Sherin, B. (2000). Meta-representation. *J. of Mathematical Behavior, 19*, 385-398.

Dove, G. (2009). Beyond perceptual symbols. *Cognition, 110*(3), 412-431.

Edwards, L., Radford, L., & Arzarello, F. (2009). Gestures and multimodality in the construction of mathematical meaning [Special Issue]. *Educational Studies of Mathematics, 70*(2).

Gelman, R., & Williams, E. (1998). Enabling constraints for cognitive development and learning. In D. Kuhn & R. Siegler (Eds.), *Cognition, perception and language* (Vol. 2, pp. 575-630). NY: Wiley.

Ginsburg, H. P. (1997). *Entering the child's mind*. NYC: Cambridge University Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Ed., 5*, 93-116.

Lakatos, I. (1976). *Proofs and refutations*. New York: Cambridge University Press.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from*. NYC: Basic Books.

Lemke, J. L. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin & R. Veel (Eds.), *Reading science* (pp. 87-113). London: Routledge.

Mariotti, M. A. (2009). Artifacts and signs after a Vygotskian perspective. *ZDM, 41*, 427-440.

Nemirovsky, R., & Ferrara, F. (2009). Mathematical imagination and embodied cognition. *ESM, 70*(2), 159-174.

Ortony, A. (Ed.). (1993). *Metaphor and thought*. Cambridge: Cambridge U. Press.

Piaget, J. (1952). *The child's conception of number*. London: Routledge.

Prawat, R. S. (1999). Dewey, Peirce, and the learning paradox. *Amer Ed Res J, 36*, 47-76.

Presmeg, N. (2006). Research on visualization in learning and teaching mathematics. In A. Gutiérrez & P. Boero (Eds.), *Handbook of research on the psych. of math. ed.* (pp. 205-235). Rotterdam: Sense.

Radford, L. (2008). Iconicity and contraction. *ZDM, 40*(1), 83-96.

Rubin, A., & Hammerman, J. (2007, Aug.). *Soup or stew: Metaphors for the relationship between samples and populations.* Paper presented at SRTL-5, Warwick, UK.

Schoenfeld, A. H. (1991). On mathematics as sense-making. In J. F. Voss, D. N. Perkins & J. Segal (Eds.), *Informal reasoning and education* (pp. vii-xvii). NJ: LEA.

Sfard, A. (2002). The interplay of intimations and implementations. *JLS, 11*(2&3), 319-357.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method. *Amer Psy, 46*(6), 606-620.

Slobin, D. I. (1996). From "thought and language" to "thinking to speaking". In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge U. Press.

Stroup, W. M. (2002). Understanding qualitative calculus. *Int J of Comp for Math Learning, 7*(2), 167-215.

Thompson, P. W. (1993). Quantitative reasoning, complexity, and additive structures. *ESM, 3*, 165-208.

Wilensky, U. (1997). What is normal anyway?: Therapy for epistemological anxiety. *ESM, 33*(2), 171-202.

Wilensky, U. (1999). *NetLogo*. Northwestern U, Evanston, IL: CCL http://ccl.northwestern.edu/netlogo/.

Wilensky, U., & Papert, S. (2009). *Restructurations*. (Manuscript in preparation).

# Quiet Captures: A Tool for Capturing the Evidence of Seamless Learning with Mobile Devices

Ivica Boticki, Hyo-Jeong So
Learning Sciences Lab, National institute of Education, Nanyang Technological University, Singapore 637616
ivica.boticki@nie.edu.sg, hyojeong.so@nie.edu.sg

**Abstract:** Seamless learning environments help students in extending their learning experiences across social, personal and environmental dimensions, and are supported by different tools, information and communication technology. In order to capture the evidence of the ways students utilize mobile devices in their formal and informal learning pursuits, a quiet data capture program was designed and pilot-tested. It is used to capture both quantitative and qualitative data about device use, to track user generated data and artefacts persistent across time and to generate and render aggregated reports identifying usage patterns. It is supposed to serve as a mechanism of identifying affordances of mobile devices within the landscape of seamless learning. Some preliminary findings raising multiple questions and inevitably sharpening our research lenses are presented towards the end of the paper.

## Introduction

Mobile devices are nowadays used by younger generations on a daily basis to access web content, to capture different artefacts (e.g. photos, voice recordings, notes etc) documenting daily activities, and to communicate with peers via synchronous (e.g. Skype for mobile devices) and asynchronous tools (e.g. SMS). However, their popularity is hardly ever extended into classrooms since school policies see mobile devices as a disruptive factor. Moreover, lesson plans and assessment strategies employed in many educational institutions usually do not allow these innovative technologies to be used as pervasive learning tools.

On the other hand, mobile devices are, according to some researchers, an enabling force to transform education (Roschelle & Pea, 2002; Sharples, Taylor, & Vavoula, 2005; Waycott, 2004; Zurita & Nussbaum, 2004). At this point, there exist numerous case studies illustrating innovative ways of incorporating mobile devices into various classroom activities (Anastopoulou, et al., 2008; Ardito, Buono, Costabile, Lanzilotti, & Pederson, 2007; Chen, Kao, Yu, & Sheu, 2004; Colella, 2000; Dufresne, Gerace, Leonard, Mestre, & Wenk, 1996; Facer, et al., 2004; Klopfer, Squire, & Jenkins, 2009; Proctor & Burton, 2003; Rogers & al., 2002; Roschelle, 2003; Sharples, Lonsdale, Meek, Rudman, & Vavoula, 2007; Vahey & Crawford, 2002). Moreover, some researchers have produced encouraging results for policy makers willing to participate in this exciting process of educational change (Klopfer, et al., 2009; Sharples, et al., 2005)

Although there is still lack of theories of mobile learning from pedagogical perspectives (Sharples, et al., 2005; Vavoula & Sharples, 2008), it is evident that mobile devices are becoming pervasively integrated into student activities outside schools. While there is no doubt that students utilize them in their daily leisure activities, it remains unclear whether there is an added educational value brought into students' lives with the help of these ubiquitously present tools. Even more intriguing questions start to emerge as we delve into this heavily under-explored area on pedagogical issues: is there a way of establishing a link between classroom and informal daily activities with the mediation of mobile devices? What kind of skills can be better taught with mobile devices? Can they be used for inquiry learning redirecting knowledge from informal into formally structured learning environments or vice versa? All these together with other emergent issues are being examined under the Seamless Learning project conducted in Singapore (see Looi, 2010 for more details). Both formal and informal learning spaces are being explored together with the synergistic effects of linking the two in order to achieve the continuity of learning experiences across different learning scenarios. The project tries to achieve so called "seamlessness" with the combinations of different parameters across multiple dimensions of learning environments.

The aim of this paper is not to answer all the above stated questions, but to focus on designing a tool that can be both used to capture and analyse invaluable information generated from everyday activities with mobile devices that primary school children are involved in. In order to collect the evidence of the ways students utilize mobile devices in their formal and informal learning pursuits, a quiet data capture program was designed and pilot-tested. The program is used to capture both quantitative and qualitative data about device ussage and to track user generated data and artefacts persistent across time and locations. The gathered data is aggregated in a form of reports helping researchers to pinpoint the areas of informal activities relevant to formal learning. Experiences gathered this way can be used in designing lesson plans which incorporate mobile devices and to accordingly design innovative applications for learning with mobile devices. Ideally, these should coincide with students' informal learning interests and broaden possibilities for seamless learning experiences.

The paper is organized as follows: we first present the theoretical background of the Seamless Learning project in relation to mobile learning. Then, we present the design and development of the quiet capture tool, followed by some collected data and preliminary reports used to better understand seamless learning experiences; and we conclude with future research plans and implications.

## From Mobile Learning to Seamless Learning

The *technological strand* of researchers in the area of mobile learning has focused on mobile devices, such as identification of device characteristics important to learning, the development of systems for mobile learning, appropriation of mobile devices for effective learning. An example of device-focused definitions of mobile learning can be described as learning activities using mobile phones, handheld computers, digital portable devices for music reproduction, digital cameras, voice recorders and digital pens (Excellence, 2006; Freitas & Levene, 2003; Wikipedia, 2009; Wood, Keen, Basu, & Robertshaw, 2003) The other strand of mobile learning research still focuses on devices, but explores the role of related technological aspects, including wireless networks, intelligent user interfaces and generally the advancements in the development of both hardware and software computer equipment which makes mobile learning possible (Boticki, Mornar, & Bozic, 2006; Holzinger, Kickmeier-Rust, & Albert, 2007; O'Malley & al., 2004; Wang, 2004). On the other hand, the *pedagogical strand* positions technological solutions coupled with the areas of learning and teaching they can best support (O'Malley & al., 2004). According to this strand, technology should be designed to support learning, rather than being appropriated to reach certain pedagogical goals. Hence, mobile devices are seen as mediating tools for activities supporting constructivist and situative approaches to learning.

Some researchers analyze the *mobility* of the learners as the primary characteristic of mobile learning (Kress & Pachler, 2007). Since learners learn on various locations, knowledge acquired at one location can be applied to others (Freitas & Levene, 2003; Sharples, 2000). Mobility of knowledge can be seen in terms of time scales as well. Previously acquired knowledge is supplemented with new ideas and strategies over time. For instance, Sharples and colleagues (2005) suggest that Some researchers suggest hat by placing the "mobility of learning as the object of analysis we may understand better how knowledge and skills can be transferred across contexts such as home and school, how learning can be managed across life transitions, and how new technologies can be designed to support a society in which people on the move increasingly try to cram learning into the interstices of daily life (p.2)".

The *Seamless Learning* project in Singapore presented in this paper examines synergistic effects of linking formal and informal learning environments in order to achieve the continuity of learning experiences across different learning scenarios. It examines both individual and social approaches to learning and tries to determine the role of one-to-one TEL (technology enhanced learning) in Primary (elementary) School children's learning experiences (Looi, et al., 2010)..

From pedagogical perspective, seamless learning examines an inevitable shift from teacher-centred to learner-centred learning environments and tackles the issues of traditional schooling practices suffering from the excessive amount of decontextualized information, indirect and abstract knowledge, and second hand experiences (Barab, 2002). Many of today's classroom learning still have a strong focus on individual cognition, pure mental activity without tool use and overly context-general learning (Brown, Collins, & Duguid, 1989). Through the notion of seamless learning, we aim to examine learning that takes place through the individual learning in private spaces, collaborative learning in public spaces, together with the cognitive artefacts created across time and physical or virtual spaces mediated by technology within a context. These artefacts facilitate knowledge construction and social discourse, and mediate interaction among a community of learners. To achieve so, we have employed distributed cognition as a theoretical lens and have integrated ethnographic approaches into our research design, especially in examining informal learning experiences (Hollan, Hutchins, & Kirsch, 2001; Looi, et al., 2010).

Our research is twofold: on the one hand, it examines formal learning environments and proactively tries to "mobilize the curriculum" or to achieve systematic changes in terms of how in-class learning can be implemented, sustained, and assessed through the use of mobile technologies. For this purpose, existing curricula have been "unpacked" to designing lesson plans that promote collaborative, inquiry-centred, and self-directed learning experiences (Looi, et al., 2009). Mobile devices and applications play an important role in this. For example, our participants, 39 primary school students, organize and externalize their knowledge with the help of GoKnow mobile learning suite (Figure 1), practice collaborative skills using the Seamless Mobile Forum (Figure 2), and other applications available on their HTC Windows Mobile 6.0 based smartphones.

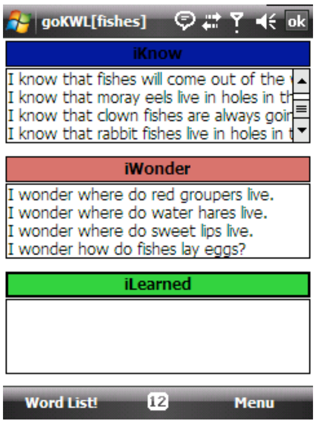



Figure 1. goKWL module of the GoKnow's MLE Mobile Learning Suite

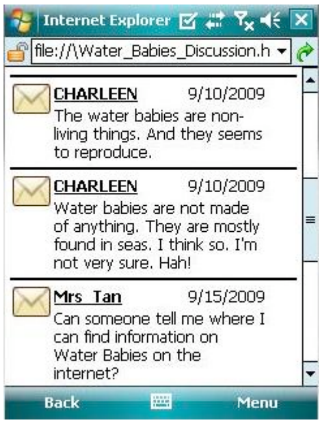



Figure 2. Seamless Mobile Forum Created Within the Seamless Learning Project

## Capturing the Evidence of Seamless Learning on Mobile Devices

### Choosing a Data Collection Approach

In addition to examining seamlessness within formal learning environments, our goal is to examine student-initiated learning that is impromptu and emergent (Sefton-Green, 2004). To achieve that, we have decided to employ a concurrent mixed method as a data collection and interpretation approach (Creswell, 2009). Quantitative data collection helps us in getting the idea of general trends while qualitative methods allow us to extract specifically significant patterns or cases for more in-depth analysis. Qualitative and quantitative data collection mechanisms are exercised concurrently, thus bringing us towards a clearer picture of informal learning spaces, together with the role of mobile devices in facilitating such learning experiences.

In our Seamless Learning project, qualitative data collection methods include sustained classroom observations, interviews with the students and their families, and observational studies (Sefton-Green, 2004). Sustained classroom observations enable us both to gain the holistic picture of the classroom atmosphere and to capture the emergent social processes. Interviews are done in participants' homes and include students and their family members primarily focusing on the informal learning experiences outside classrooms. Observational studies are used in designed spaces (currently in a science museum) to capture both video and sound clips of entire family visits to examine family interaction and discourse in a designed informal learning space.

Quantitative instruments utilized in the project include surveys, experiments and log files. Surveys primarily serve as a mechanism for longitudinal evaluation used to compare formal assessment results and students' attitudes towards the mobile device use at project milestones. Survey instruments therefore both guide us in our future research pursuits and provide us with a means of evaluating long-term effects of our inevitably interventionist approach. Experiments are used to test out novel tools, such as mobile computer supported fractions software, primarily in order to gather usability and learning experience feedback. Unobtrusive methods such as log files provide an authentic, time-efficient means of recording student learning behaviours and can capture a wide range of data that reflect student practices, activities, context, situations and events (Buckley, Gobert, & Horwitz, 2006; Cole, 1995).

### Mixed Method Approach to Quiet Captures

One of the common methods of collecting data from mobile devices is through log files that support an objective data collection method of acquiring quantitative data on device usage. Due to the convergence trend (i.e., mobile and communication devices merging into a single device), mobile devices are becoming powerfully connected computers allowing the development of specialized applications to be installed and used on them. Log files are sometimes referred to as the quiet captures emphasizing the unobtrusiveness of the way they are embedded into user's experiences. They typically capture the data on applications used and the overall time spent.

In addition to their obviously limited coverage and purely quantitative orientation, current log file applications require researchers to periodically gather logs from participants' devices, thus to a certain level interfering their experiences. Furthermore, user generated artefacts are not tracked persistently across time (i.e. students delete self-created videos), and do not capture the context of student-device interaction during the identified patterns of usage. *Therefore we propose a mixed mode quiet capture approach with the following features: (a) the ability to capture both quantitative and qualitative data about device usage, (b) the ability to transfer and track user generated data and artefacts persistently (i.e., created movies, sound clips and pictures) to researchers in real-time 24/7 and (c) the ability to generate and render aggregated reports identifying usage patterns.*

Qualitative approaches to collecting device usage data consist of capturing device screenshots when some specific applications are used and/or when users start interacting with the device. Those screenshots are timely ordered and sent to the central repository through the 3G wireless mobile connection. Captured screens are then available to researchers in form of a web site and can be browsed and filtered according to the attached contextual information. Quantitative approaches to collecting data rely on the classical log file method of collecting the duration of application usage extended with a possibility of transferring content to the central repository in real-time.

Our subjects are equipped with mobile broadband plans with unlimited data transfer package. In this context, 3G connection serves as a means of providing affordable real-time 24/7 data transferring channels to central data repositories. This eliminates two drawbacks of the classical quiet captures approaches: (a) researchers do not have to manually gather the collected data from mobile devices in order to analyze it, and (b) data is available for the analysis whenever the need exists. Moreover, student generated artefacts are stored across time: pictures, files and videos are stored and become inerasable proof of user activities available for future analysis.

Researchers are provided with the web interface through which qualitative and quantitative data can be analyzed (Figure 3 on the next page). Quantitative data can be examined through the various reports while qualitative data is given to researchers for further detailed analysis. Both can serve as a connecting point not only in cross-modality verification of the results, but also in the interpretation of data acquired by employing various collection methods.

## Quiet Capture Tool Architecture

The tool gathers various types of information, such as application usage, student generated artefacts and screenshots to transmit them to the server side web service over the 3G wireless network. The transferred data is kept persistently into the server-side database for later retrieval. The stored data is available to researchers in the forms of predesigned reports 24/7. Students can therefore be monitored real-time, and usage statistics are generated on demand (Figure 4.)

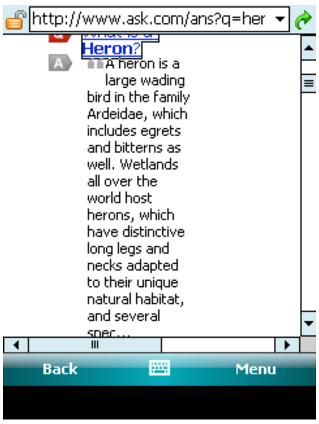



Figure 3. A (quiet) screen capture used to examine emergent inquiry based learning (a student searching for information on Heron birds)

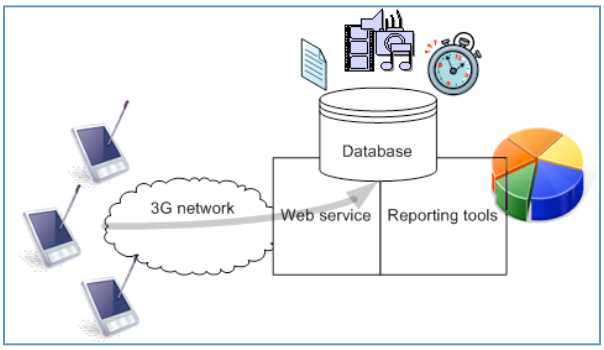



Figure 4. Quit capture tool system architecture

Needless to say, such a design requires mobile devices to be equipped with mobile broadband wireless Internet access plans with adequate bandwidth. In our project, students, in addition to the devices themselves, are given the 24/7 Internet access with unlimited data transfer plans. Instances of the quiet capture tool were installed onto the student devices, and they run as long as the devices are on. As soon as students decide to turn off the device, the tool's activity is suspended, thus saving resources such as battery life and Internet data connection. The client-side component of the quiet capture tool is composed of the following modules: Activity Detection Module, Artefact Gathering Module, Screen Capture Module and Data Transfer Module.

## Reports on the Formal and Informal Device Use

The value of the described quiet capture approach lies in the analysis performed on the collected data generating aggregated reports about device use. At this point, we are able to share rough reports generated from the data of our 39 primary school students tracked both in and out school over one month. Although the amount of data is not significant to draw any final conclusions, we feel that the reports match and supplement the findings from other employed data collection instruments and surely inform us of behaviours exhibited by the students.

To depict the average time students spent on using the devices, we employ the functional framework in informal learning activities on mobile platforms (Clough, Jones, McAndrew, & Scanlon, 2007; Patten, Sánchez,

& Tangney, 2006). Activities are therefore classified into (a) Collaborative, Location-aware, (b) Data collection, (c) Microworld, (d) Administration, (e) Referential, (f) Appropriation and (g) Interactive (Figure 5).

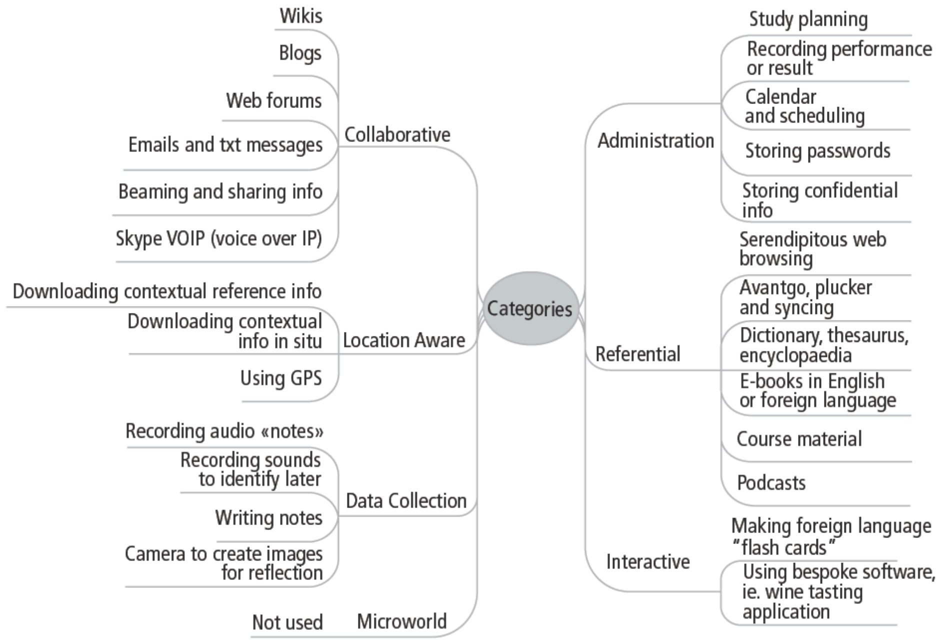



Figure 5. The map of the functional Framework in informal learning activities on mobile platforms (Clough, et al., 2007; Patten, et al., 2006)

Figure 6. shows that our students were mostly involved into the "Referential" activities (46% of the average time a student spent on purposely using the device) which include web browsing, checking dictionaries, accessing course materials, etc. In addition to that, students were often engaged into the Data Collection, Interactive Activities and Games. Since Patten et al's (2006) functional framework does not clearly state which category should mobile computer games assigned to, in our analysis we are considering them as a separate category (Figure 6).

To get a clearer picture on the data presented in the Figure 6, a more detailed report is made available to researchers (Table 1). In addition to the average time students spent on a specific activity type, the report shows minimum, maximum values and the standard derivation extracted from a month data of student activities. This report makes the differences in student interest more evident, particularly when some categories, such as games or interactive applications are concerned.

In order to gain an in-depth picture on the specific application use, researchers can drill-down the report and obtain specific application usage statistics (Figure 7). The figure shows 15 applications students on average spent the most time on (in minutes). The analysis reveals that Web browsing applications such as the Internet Explorer and Skyfire are the most utilized tools on students' devices, closely followed by the applications for watching videos such as YouTube, Player, and Streaming Media.

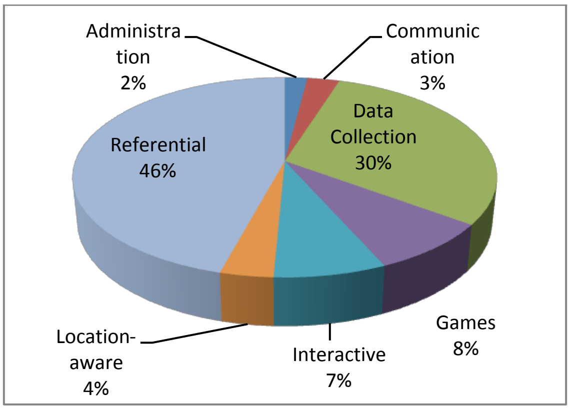



Figure 6. Distribution of student device use time according to the application types

Table 1. Monthly time (in minutes) students spent on an application type

| *Application type* | *MIN* | *MAX* | *AVG* | *STDEV* |
|---|---|---|---|---|
| **Administration** | 0.02 | 47.55 | 4.72 | 8.78 |
| **Communication** | 0.03 | 40.95 | 6.77 | 10.74 |
| **Data Collection** | 24.95 | 150.58 | 70.93 | 0.28 |
| **Games** | 0.12 | 169.13 | 19.43 | 3.55 |
| **Interactive** | 5.50 | 65.68 | 17.27 | 13.56 |
| **Location-aware** | 0.05 | 29.95 | 8.18 | 10.13 |
| **Referential** | 2.67 | 705.42 | 107.58 | 12.58 |

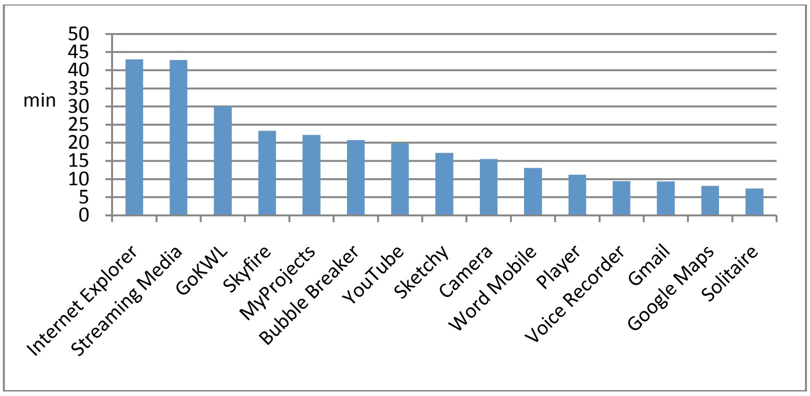



Figure 7. Average monthly time (in minutes) a student spent on using specific application (given only for the 15 top applications)

One of the main aims of our project is to examine the relationship of the formal and informal learning environments. In order to achieve so, we have analyzed specific application usage according to the time of the day: Table 2 has specific application names and times of the day as its two-dimensional base, while the third dimension contains average student monthly usage. The chart serves as a visual aid in determining the patches of significant user activities enabling researchers to drill down and identify specific sub-categories of interest.

Table 2. A table showing average monthly time (in seconds) a student spent on a specific application according to the time of the day

| Application name/Time of the day | 0 | 1 | 2 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Skyfire | | | | 279 | 1635 | 1699 | 1557 | 33 | 184 | 328 | 359 | 307 | | 1466 | 73 | 474 | 995 | 2 | 2766 | 728 | 391 | 15 | |
| Streaming Media | 321 | | | | | 301 | 383 | 564 | 565 | 346 | 695 | 1019 | 345 | 774 | 898 | 557 | 634 | 349 | 374 | 414 | 1124 | 443 | 485 |
| Internet Explorer | 1924 | | 510 | | | 352 | 141 | 568 | 328 | 650 | 548 | 367 | 301 | 294 | 373 | 373 | 221 | 350 | 235 | 123 | 358 | 198 | 1620 |
| Youtube | 281 | | | | | | | 442 | 5 | 1286 | 35 | 339 | 297 | 1378 | 984 | 674 | | 436 | | 5 | 2 | 116 | 688 |
| Bubble Breaker | | | | | | 81 | | 105 | 296 | 201 | 162 | 1936 | 261 | 205 | 507 | 660 | 934 | 693 | 147 | 238 | 212 | 45 | |
| Sketchy | | | | | | | | 284 | 110 | 109 | 740 | | 105 | 540 | 315 | 1265 | | | 2566 | 15 | | 454 | 1 |
| Google Maps | 211 | | | | | | | 988 | | 3 | 13 | | | 997 | | 59 | | 26 | | 800 | 166 | 628 | 61 |
| Camera | | | | | | | 189 | 79 | 24 | 104 | 531 | 240 | 144 | 15 | 217 | 727 | 362 | 18 | | 146 | 4 | 2 | |
| Voice Recorder | | | | | | | | 107 | 4 | 3 | 131 | 118 | 5 | | | | 531 | 656 | 823 | | 196 | 41 | |
| Player | | | | | | | | 92 | | 33 | 117 | 122 | 17 | 17 | 419 | 419 | 142 | 214 | 157 | 186 | 625 | | |
| Solitaire | | | | | | | | 19 | 14 | | 78 | | | | 471 | 1204 | 71 | 106 | | | | 91 | |
| Gmail | | | | | | 101 | | 148 | 237 | 279 | 292 | 59 | 83 | 313 | 68 | 73 | 72 | | 56 | 21 | 80 | 21 | 12 |
| GoKWL (GoKnow) | | | | | | | | | 5 | 699 | 1189 | | | | | 16 | | | | | | | |
| MyProjects | | | | | | | | 124 | 105 | 599 | 576 | 197 | | 51 | 26 | 86 | | | 14 | 38 | 15 | 59 | |
| Word Mobile | | | | | | | | 34 | 22 | 35 | 667 | 267 | | | 94 | 6 | | | | | 79 | 89 | |
| Grand Total (s) | 2737 | 0 | 510 | 279 | 1635 | 2534 | 2270 | 3587 | 1899 | 4675 | 6133 | 4971 | 1558 | 6050 | 4445 | 6593 | 3962 | 2850 | 7138 | 2714 | 3252 | 2202 | 2867 |

Researchers might opt to examine a specific application for a more refined analysis. For example, the contour chart signals high overall usage of the Internet Explorer application throughout the entire day, especially being the extreme in the end of the day.

## Conclusion and Future Plans

The paper presented one of the tools in the mixed mode approach for capturing mobile device use data from both formal and informal learning environments. We presented the quiet capture tool to gather application usage information, student-created artefacts and screenshots. Researchers can use the reports generated from the collected data to examine students' use of mobile devices across different location and time scales.

Preliminary findings presented in this paper show that students tend to spend a significant amount of time on using the device in their personal spaces. They mostly browse the Internet for information and videos, and create interactive digital artefacts. For some students, the informal use of mobile devices produces a shift in cognitive processing skills, while for the others the device has become a social mediating tool facilitating interaction with the classmates.

It should be noted that all data reported in this paper is collected with the participants' consensus for confidentiality and privacy issues. Researchers need to consider that such quiet capture tools may collect private data unwanted by the participants so that it is important to ensure user anonymity and ethical use of collected data.

Future research plans include working with researchers on developing new views on the collected data, developing triangulation methods for merging data gathered from other data collection methods and further expanding the amount of data to be collected. As an example, there are plans to introduce geo-tagging of all collected information in order to enable the creation of location-based reports. In addition to analysing informal learning experiences according to a new spatial dimension, we will employ data mining and aggregation techniques to determine in what extent our students use the devices in and out school, specifically focusing on several spatio-temporal categories: in-school, holiday, during weekends and during the week.

Although at this point we observe that students use mobile devices to learn beyond the walls of classroom and formal institutions, and the usage clearly impacts their formal learning experiences, we are still in the process of designing mechanisms to clearly prove so. We believe the Quiet Captures tool presented in this paper can be used to give us better understanding of the context in which the links emerge and vanish. We have only begun to understand the complex interplay of formal and informal learning spaces constantly having in mind that it is our job to nurture the link between formal and informal learning environments and to utilize this unique opportunity to promote holistic learning experiences.

## Acknowledgements

This study is supported by a grant from the National Research Foundation, Singapore. We are grateful to the researchers from the Learning Sciences Laboratory (LSL), National Institute of Education (NIE), Nanyang Technological University (NTU), Singapore collaborating with us on this research.

## References


Anastopoulou, S., Sharples, M., Wright, M., Martin, H., Ainsworth, S., Benford, S., et al. (2008). *Learning 21st century science in context with mobile technologies*. Paper presented at the mLearn 2008 Conference.

Ardito, C., Buono, P., Costabile, M. F., Lanzilotti, R., & Pederson, T. (2007). *Re-experiencing history in archaeological parks by Playing a Mobile Augmented Reality Game*. Paper presented at the OTM MONET'07 Workshop, Springer LNCS 4805.

Barab, S. (2002). Human-field interaction as mediated by mobile computers *CSCL 2: carrying forward the conversation* (pp. 533–537). Mahwah, NJ: Lawrence Ealbaum.

Boticki, I., Mornar, V., & Bozic, N. H. (2006). *Introducing Location - Awareness into a Learning Environment Supported by Mobile Devices*. Paper presented at the 1st International Conference on Virtual Learning.

Brown, J. S., Collins, A., & Duguid, S. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Buckley, B. C., Gobert, J. D., & Horwitz, P. (2006). *Using log files to track students' model-based inquiry*. Paper presented at the International Conference of the Learning Sciences.

Chen, Y. S., Kao, T. C., Yu, G. J., & Sheu, J. P. (2004). *A mobile butterfly-watching learning system for supporting independent learning*. Paper presented at the 2nd International Workshop on Wireless and Mobile Technologies in Education.

Clough, G., Jones, A. C., McAndrew, P., & Scanlon, E. (2007). Informal learning with PDAs and smartphones. *Journal of Computer Assisted Learning, 24*(5), 359–371.

Cole, M. (1995). The supra-individual envelope of development: activity and practice, situation and context *Cultural practices as contexts for development*. San Francisco: Jossey-Bass.

Colella, V. (2000). Participatory simulations: building collaborative understanding through immersive dynamic modeling. *Journal of the Learning Sciences, 9*(4), 471-500.

Creswell, J. W. (2009). *Research design: quantitative, qualitative and mixed-methods approaches 3rd ed.*: Sage.

Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., & Wenk, L. (1996). Classtalk: a classroom communication system for active learning. *Journal of Computing in Higher Education, 7*(2), 3-47.

Excellence, K. N. o. (2006). *Mobile Learning Initiative Big Issues in Mobile Learning*: University of Nottingham, United Kingdom.

Facer, K., Joiner, R., Stanton, D., Reid, J., Hull, R., & D.Kirks. (2004). Savannah: mobile gaming and learning? *Journal of Computer Assisted Learning, 20*(6), 399-409

Freitas, S. D., & Levene, M. (2003). *Evaluating the development of wearable devices, personal data assistants and the use of other mobile devices in further and higher education institutions* (Tech. Rep.). United Kindgom: JISC.

Hollan, J., Hutchins, E., & Kirsch, D. (2001). Distributed cognition: toward a new foundation for human-computer interaction research *Human computer interaction in the new millennium* (pp. 75–94). New York: ACM Press.

Holzinger, A., Kickmeier-Rust, M., & Albert, D. (2007). Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? *Educational Technology & Society, 11*(1), 279-290.

Klopfer, E., Squire, K., & Jenkins, H. (2009). Environmental Detectives: the development o fan augmented reality platform for environmental simulations. *Education & Research Development, (in print).*

Kress, G., & Pachler, N. (2007). *Thinking about the 'm' in m-learning, in Mobile Learning towards a research agenda*: Institute of Education, London.

Looi, C. K., Seow, P., Zhang, B., So, H. J., Chen, W.-L., & Wong, L. H. (2010). Leveraging mobile technology for sustainable seamless learning. *British Journal of Educational Technology, 42*(1).

Looi, C. K., Wong, L. H., So, H. J., Seow, P., Toh, Y., Chen, W., et al. (2009). Anatomy of a mobilized lesson: Learning my way. *Computers & Education, (in press).*

O'Malley, C., & al., e. (2004). *WP4 – Pedagogical Methodologies and Paradigms, Guidelines for learning/teaching/tutoring in a mobile environment*: MOBIlearn.

Patten, B., Sánchez, I. A., & Tangney, B. (2006). Designing collaborative, constructionist and contextual applications for handheld devices. *Computers & Education, 46*(3), 294 - 308.

Proctor, N., & Burton, J. (2003). *Tate Modern multimedia tour pilots 2002-2003*. Paper presented at the MLEARN 2003: Learning with Mobile Devices.

Rogers, Y., & al., e. (2002). *Learning through digitally-augmented physical experiences: reflections on the Ambient Wood project* (Technical Report): Equator

Roschelle, J. (2003). Keynote paper: Unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning, 19*(3), 269-272.

Roschelle, J., & Pea, R. (2002). A walk on the WILD side: How wireless handhelds may change computer-supported collaborative learning. *International Journal of Cognition and Technology, 1*(1), 145-168.

Sefton-Green, J. (2004). Literature review in informal learning with technology outside school [Electronic Version]. Retrieved from http://www.futurelab.org.uk/resources/documents/lit_reviews/Informal_Learning_Review.pdf

Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers and Education, 34*(3), 177-193.

Sharples, M., Lonsdale, P., Meek, J., Rudman, P. D., & Vavoula, G. N. (2007). *An Evaluation of MyArtSpace: a Mobile Learning Service for School Museum Trips*. Paper presented at the 6th Annual Conference on Mobile Learning, mLearn 2007.

Sharples, M., Taylor, J., & Vavoula, G. (2005). *Towards a theory of mobile learning*. Paper presented at the MLEARN2005 Conference.

Vahey, P., & Crawford, V. (2002). *Palm Education Pioneers Program* (No. Final Evaluation Report). Menlo Park, CA: SRI International.

Vavoula, G., & Sharples, M. (2008). *Challenges in Evaluating Mobile Learning*. Paper presented at the mLearn 2008 Conference.

Wang, Y.-K. (2004). *Context Awareness and Adaptation in Mobile Learning*. Paper presented at the 2nd IEEE International Workshop on Wireless and Mobile Technologies (WMTE'04).

Waycott, J. (2004). *The appropriation of PDAs as Learning and Workplace Tools: An Activity Theory Perspective.* The Open University, Milton Keynes, UK.

Wikipedia. (2009). Mobile Learning. Retrieved 2 Nov 2009, 2009, from http://en.wikipedia.org/wiki/M-learning

Wood, J., Keen, A., Basu, N., & Robertshaw, S. (2003). *The Development of Mobile Applications for Patient Education, in Designing for User Experiences (DUX).* San Francisco, USA.

Zurita, G., & Nussbaum, M. (2004). Computer supported collaborative learning using wirelessly interconnected handheld computers. *Computers & Education, 42*(3), 289-314.

# The Many Dimensions of Having a Good Eye: A Methodological Reflection of Metaphors in Visual Cognition Analysis

Andreas Gegenfurtner, Anna Siewiorek, Centre for Learning Research, University of Turku, Finland
Email: angege@utu.fi, amsiew@utu.fi

**Abstract:** There seems to be a stable and widely held belief among medical practitioners that good diagnosticians have a good eye that is innate, rather than trained. However, there is no evidence for any identifiable perceptual trait. Nonetheless, the process of developing a good eye in medicine is proposed, indicated, and elaborated by various measures contingent on diverse methodological arenas, all of which attempt to advance our understanding of what constitutes visual cognition in diagnosing medical images. The purpose of this paper is to provide a reflection on this methodological pluralism. We first identify four metaphors used in the analysis of visual cognition: activation, detection, inference, and practice. These metaphors are described with an empirical example and discussed to elicit (partly tacit) assumptions associated with prototypical method decisions. We then link the proposed metaphorical mapping to what it implies for current epistemological, methodological, and curricular discussions in medical education.

## Introduction

There seems to be a stable and widely held belief among medical practitioners that good diagnosticians are born rather than made (Elstein, Shulman, & Sprafka, 1978; Ericsson, 2004; Norman, Coblentz, Brooks, & Babcook, 1992). Indeed, it is a common observation, both in training programs and in clinical practice, that some residents are quicker and more accurate than others in recognizing subtle abnormalities in medical images. Based on everyday observation of these kinds, it has subsequently been concluded that diagnostic accuracy is due to having a "good eye"—one that is innate, rather than trained. However, there is no evidence for any identifiable perceptual trait. To the contrary, recent research has indicated that what has been labeled as a good eye can in fact be developed through training. For example, Dawes, Vowler, Allen, and Dixon (2004) demonstrated that detection performance of clinical students was significantly higher after a 26-week training period than it had been before; materials included graphics used in diverse medical disciplines such as plain X-ray films, ultrasound, computer tomograms, magnetic resonance images, and positron emission tomograms. Similar training effects have also been reported in cardiology (Issenberg, McGaghie, Gordon, Symes, Petrusa, Hart, et al., 2002), radiology (Brazeau-Lamontagne, Charlin, Gagnon, Samson, & Van der Vleuten, 2004), and microscopic pathology (Crowley, Lecowski, Medvedeva, Tseytlin, Roh, & Jukic, 2006). Overall, we can conclude that training can improve visual accuracy; that recognition of abnormities in perceptual tasks can be learned; and that good diagnosticians are made rather than born.

Thus, there is consensus that the development of being a good diagnostician goes in concert with the development of a good eye. What is less clear is what a good eye is. What precisely is it that develops in novice diagnosticians through medical education? What must be appropriated to move toward higher accuracy? Some investigators, particularly Kundel, Nodine, and Carmody (1978), in a now classic study, suggest that the highly perceptual nature of image comprehension requires intensive processing of visual data through oculomotor activity that guides signal detection and decision-making. Others, particularly Lesgold and colleagues (1988), in what is now also a classic study, put less emphasis on the perceptual aspect; rather, they suggest that diagnosing is mainly a cognitive inference that aligns disease schemata from episodic memory consistent with the perceptual features detected. Much medical research done in both traditions has been reviewed elsewhere (Boshuizen & Schmidt, 2008; Ericsson, 2004; Kundel, 2007; Norman et al., 1992; Patel, Arocha, & Zhang, 2005). Recently, alternative approaches have been formulated; these suggest that a good eye is indicated by neurophysiologic events in certain brain areas (Haller & Radue, 2005), and is accomplished through situated social discourse (Koschmann, LeBaron, Goodwin, Zemel, & Dunnington, 2007). In short, the metaphor of having a good eye in medicine is proposed, indicated, and elaborated by various measures contingent on diverse methodological arenas that all attempt to advance our understanding of what constitutes visual cognition.

The main task of the present study is to provide a first step in proposing a reflection on this methodological pluralism. Our goal is to present a framework in which the many dimensions of having a good eye can be considered as mutually constituting the richness we have on ideas, concepts, and theories relating to medical image diagnosis. We have no intent to judge some methodological traditions as being more valuable than others, nor do we intend to unify them in some abstract way. Although the idea of what was termed *interactive complexification* (Alexander, Schallert, & Reynolds, 2009) is considered meaningful for highlighting the confluence of factors that determines "any given aspect" of the product and the process of learning to diagnose medical images, we believe there is, at times, also a place for simple answers. We hope that the framework put forward in this study will provide such a simple answer. The study addresses the questions of how to analyze visual cognition; how to elicit tacit assumptions underlying common research practice; and how to inform curriculum design for medical training in corporate and higher education settings. By discussing four examples drawn from radiology, we identify four metaphors that constitute our framework on visual cognition analysis. Despite its groundwork in medical literature, it could be assumed that the framework also has implications beyond medicine, due to its common interest in learning and comprehension in technology-rich vision-intensive contexts. In the first part of this paper, we introduce four metaphors for visual cognition: activation, detection, inference, and practice. In the second part, we discuss the potential implications of these metaphors in reconsidering standpoints on epistemology, methodology, and curriculum design.

## Four Dimensions of Visual Cognition Analysis: A Metaphorical Mapping

Metaphors are a useful tool for mirroring in simple terms the often complex fundamentality underlying theory and applied research practice. An example of how metaphors are able to elicit complex, different, yet partly tacit, assumptions on a commonly studied phenomenon is the reflection on learning by use of the metaphors acquisition, participation, and knowledge creation (Paavola, Lipponen, & Hakkarainen, 2002; Sfard, 1998). As Sfard (1998, p. 4) notes, "metaphors are the most primitive, most elusive, and yet amazingly informative objects of analysis". We believe that their value and power stems from the fact that metaphors converge and portray, in a snapshot format, what took years of scientific discourse to develop; this allows frank presentation of positions and their entailments in a condensed way and invites a critical (re)consideration of accepted and perhaps unreflected practice. Of course, metaphors are simple and simplistic; there is no claim that they attempt to depict all of the breadth and depth of what often is a complex epistemology.

Below, we identify, exemplify, and discuss four metaphors often used when analyzing visual cognition; these metaphors are seen as four of the many dimensions possessed by one with "a good eye" in medicine. We reflect on these metaphors in terms of what they contribute as an answer for the two questions stated above: The nature of what develops in novice diagnosticians through medical education is not yet clear. The question of what must be appropriated to move toward higher accuracy remains to be answered. It is hoped that the value and significance of this methodological reflection will be that it helps to map a scattered and fragmented field of inquiry. Table 1 serves as the guiding framework for the comparison of the four metaphors, including their methodological entailments: activation, detection, inference, and practice.

Table 1: A comparative metaphorical mapping of visual cognition analysis in medicine.

| | Activation | Detection | Inference | Practice |
|---|---|---|---|---|
| Indicators of visual cognition | Neurophysiologic activity | Eye movements | Verbal reports | Representational practices |
| Unit of analysis | Individual | Individual (social) | Individual and social | Sociotechnical |
| Place of visual cognition | Neural network system | Optic system | (Distributed) memory system | Activity system |
| Analytic time span | Milliseconds | Seconds | Minutes to few hours | Minutes to decades |
| Associated methodology | Cognitive neuroscience | ROC analysis; eye tracking methodology | Protocol analysis | Ethnomethodology |

### Activation Metaphor

Research adhering to the activation metaphor uses measures of neurophysiologic activity as an indication of visual cognition. In the activation metaphor, there is a strong emphasis on the neurological and biological basis of our *humanness* (Alexander et al., 2009; see also Meltzoff, Kuhl, Movellan, & Sejnowski, 2009). This emphasis might originate from the widely held belief that "information is stored in neural networks in the brain, and that human behavior arises from extremely complex communication between neurons in these networks and also between separate networks or assemblies" (Sauseng & Klimesch, 2008, p. 1003). This neural network system is seen as the place where visual cognition occurs. Hence, visual cognition is indicated by neural activity. Specifically, this activity can be measured by an electroencephalograph (EEG) as the electric current in axons; by a magnetoencephalograph (MEG) as the magnetic field induced by those electric currents; by a positron emission tomography (PET) as the blood flow distribution in the cells; or by functional magnetic resonance imaging (fMRI) scanning as cellular oxygen consumption. Whenever information stored in neural networks is used for cognitive processes, neural activation occurs that can be measured with one of those tools. For example, if a radiologist formulates a diagnosis based on a patient's fMRI scan, an fMRI scanner could be used to indicate the processes of this radiologist's visual cognition. These processes are extremely fast; the best conventional apparatus currently available—the EEG scanner—is able to trace this activity with a temporal resolution in the range of milliseconds (Sauseng & Klimesch, 2008). An empirical example prototypical for the activation metaphor can further illustrate epistemological and methodological premises.

Haller and Radue (2005) investigated differences in neuronal activations of radiologists and laypersons in reading radiologic and non-radiologic images. Using functional magnetic resonance (fMRI) imaging, the brain scans showed that radiologic images evoked stronger activations in the brains of radiologists than in those of laypersons, with the bilateral middle and inferior temporal gyrus, bilateral medial and middle frontal gyrus, and left superior and inferior frontal gyrus being particularly affected. These regions are generally assumed to be linked to the encoding and storing of memory of visual objects and events. Hence, this finding seems to imply that what is seen on the presented image is automatically referenced to memorized images, indicating an unconscious, stimulus-driven indexical relation between the pictorial representation and the corresponding mental representation. Being prototypical for research in neuroscience, Haller and Radue (2005) used technological images as stimuli in a series. Stimuli were shown for 2.5 seconds followed by a fixation cross for 8.5 seconds to compensate for blood oxygenation level-dependent signal delay. Subjects gazed at the stimuli series with an immobilized head in darkened and (electrical and auditory) noise-protected rooms. Settings of this kind are highly controlled. These types of controls are necessary because neural measures are highly sensitive: activation should be shown in response to the stimulus only. Strong controls therefore aim to guarantee bias-free recordings.

At this point in time, the activation metaphor for visual cognition has rarely been used in medical diagnosis studies. However, the coming together of the learning sciences and neuroscience is beginning to form an exciting new field (Ansari & Coch, 2006; De Jong, Van Gog, Jenks, Manlove, Van Hell, Jolles et al., 2009). An answer comes into play regarding the question of what develops in novice diagnosticians through medical education; namely, that neuroscience has the potential to trace implicit and experiential learning before it can be observed in behavior. This will allow us to understand when, how, and why learning occurs. In particular, the how and why can be tackled within these highly controlled settings. Certainly, it is not a novel statement that behaviors, such as diagnosing a medical image, that appear similar on the surface may involve very different cognitive/perceptual mechanisms underlying this behavior. Neuroscience, in combination with the learning sciences, now provides a new avenue for tackling these issues, to further understand expertise development in novice diagnosticians. Results derived from the activation metaphor therefore have the potential to inform the design of medical education and training; more on this issue will be offered at a later point in this paper.

## Detection Metaphor

Detection can be defined as "determining whether a simple, featurally defined stimulus is present in, or absent from, the visual field" (Smith & Ratcliff, 2009, p. 283). A central premise of research using the detection metaphor is uncertainty, that is, the degree to which a subject is able to discriminate between signal (the stimulus of interest) and noise (background stimuli distracting visual attention, thus causing decision-making under conditions of uncertainty). In medical image diagnosis, the signal would be a tumor, while noise would be (healthy) organic material surrounding the tumor. Clearly, in pictures as complex as radiographs, with an abundance of structures, forms, and elements displayed in manifold shadings of grey, detection of a tumor is a challenging task. Tasks of this kind are frequently used to quantify the ability of discerning between signal and noise. Two approaches are prevalent: eye-tracking methodology and receiver operating characteristic (ROC) analysis. Below, an empirical example by Kundel, Nodine, Conant, and Weinstein (2007), which combined eye-tracking with ROC analysis, can illuminate prototypical premises of each approach.

Kundel et al. (2007) investigated rapid initial fixations (detections) on abnormalities on mammograms. Briefly, they found that more experienced radiologists showed global perceptual processes that helped them detect the abnormality (malignant breast lesions) in less than a second. In contrast, less experienced radiologists showed search-to-find strategies that took considerably longer to first fixate the abnormality. Visual cognition differences in these two groups were indicated by eye-tracking and by ROC analysis. With respect to eye-tracking, the recording of eye movements is usually used to visualize the scan paths of observers. In Kundel's study, radiologists with more experience had longer saccades and fewer fixations than did less experienced radiologists. This empirical evidence is consistent with many studies that adhere to the detection metaphor in radiology (Mello-Thoms, 2003), cardiology (Augustyniak & Tadeusiewicz, 2006), and microscopic pathology (Krupinski et al., 2006). With respect to ROC analysis, detectability was significantly higher for observers with more experience than for those with less experience. Detectability is a measure that quantifies the sum of true positives and true negatives, divided by the sum of all positives and negatives in a detectability value, $d_a$.

A conclusion can be made that the detection metaphor indicates that novice diagnosticians develop a good eye through medical education in terms of their ability to discriminate a potential signal from background noise. This ability can be quantified and expressed mathematically in a formula that allows comparison of observers at the individual or the group level. Studies in the detection metaphor, which usually employ eye-tracking methodology and/or ROC analysis, indicate that superior visual cognition can be characterized as a high decision-speed accuracy relation: Visual perception changes with a rise in experience, from a relatively slow search-to-find mode to a global holistic mode. This change then increases sensitivity, specificity, and accuracy of the detection performance. Usually, the analytic time span is longer than the time span of cognitive neuroscience studies. The work of Kundel et al. (2007), which can be seen as a prototypical example, reported an average search time of 26.90 seconds, and a median time to first fixate the abnormality (the signal) of 1.13 seconds. Traditionally, eye-tracking studies have focused on individual observers as the unit of analysis; however, recent developments of eye-tracking technology and of new analytic algorithms now allow collaborative gaze studies (e.g., Sangin, Molinari, Nüssli, & Dillenbourg, 2008). It will be fascinating, from an epistemological point of view, to follow the coping of tension between attentional detection as an individual quantifiable performance, notable in mathematical functions (i.e., Smith & Ratcliff, 2009) on one hand, and detection as collaborative achievement and intersubjective meaning-making (much in line with Koschmann & Zemel's, 2009, notion of *discovery as occasioned production*), on the other hand.

## The Inference Metaphor

Lesgold and colleagues (1988, p. 336) speculated that radiological diagnosis "is largely a matter of cognitive inference. That is, given a set of findings (perceptual features), one has to determine which diseases are consistent with those findings. If more than one disease is consistent, then one either looks further, (…) or suggests additional medical tests to discriminate among the possibilities". Two issues are striking in this initial quote. First, Lesgold emphasizes cognition and memory processes in diagnosing medical images. Back in the 1980s, this was not customary in the medical literature. Although there have been pioneering studies on cognitive processes (e.g., Patel & Groen, 1986), most focused on perceptual processes (based on Arnheim, 1969, see also the section on the detection metaphor). What Lesgold indicated and empirically tested was thinking as an essential function in medical diagnosis. Second, in this quote and elsewhere in his chapter, Lesgold emphasized how vision and cognition correlated in the formation and evaluation of diagnostic decisions. Experienced radiologists build mental representations that guide perception. The literature now knows a variety of rhetorical functions built from verbal protocols to describe those mental representations, among them Lesgold's schemata, encapsulated scripts (Boshuizen & Schmidt, 2008), E-MOPs (Kolodner, 1983), SUSTAINs (Love, Medin, & Gureckis, 2004), or PANDEMONIUM processors (Selfridge, 1959). In the next paragraph, we describe a more recent example on the correlation of vision and cognition in medical image diagnosis that nonetheless is still informed by Lesgold's (one might tend to say: seminal) speculation of cognitive inference.

Morita and colleagues (2008) investigated how perceptual and conceptual processing interrelates in the diagnosis of computer tomograms (CT). Shortly, they found that expert compared to novice CT readers verbalized more findings, more hypotheses, and more perceptual activities. Importantly, experts verbalized many perceptual features during conceptual activities, and verbalized conceptual words during perceptual processing. Put differently, this indicates that experts retrieved and used knowledge from memory based on information that they saw on the CT image, which iteratively stimulated looking at the image based on knowledge coded in memory (be it in the form of encapsulated scripts, E-MOPs, or schemata). From a methodological point of view, it would be tempting to criticize the neglect to use eye movement recordings; this would have allowed highly specific, quantifiable measures of perceptual activity. Yet, verbal protocols can also be used as indicators for visual cognition. Usually, as prototypically shown in this example, protocols are collected for a duration of up to few hours and then are analyzed with a focus on cognitive mechanisms. From this perspective, the inference metaphor on visual cognition clearly emphasizes the cognitive parts of the interrelated process. Morita and colleagues (2008) decided on individual CT readers as a unit of analysis. However, protocols can also be used at a group level (Greeno, 2006; see Simpson & Gilhooly, 1997, for an example in cardiology) to indicate collaborative negotiations and intersubjective meaning-making.

The inference metaphor on visual cognition can answer, in two respects, the question regarding what develops in novice diagnosticians that moves them toward higher accuracy. First, knowledge develops; an extensive knowledge base is the foundation for expert performance and for rapid inference of coded memory to detected visible features. Second, the perceptual-conceptual processing linkage develops. Morita and colleagues have demonstrated that protocol measures are a valid tool for eliciting cognitive mechanisms underlying CT diagnoses that guide, and are guided by, perceptual detection. Epistemologically, the inference metaphor appears to account for both signal detection and the alignment of knowledge from memory (inference) that is consistent to what is detected. Nevertheless, methodologically, it lacks the precise and time-sensitive measures such as eye-tracking gaze recordings or cortical oscillation EEGs. This is because researchers rely on explicit, conscious think-aloud utterances from subjects and these cannot account for the underlying implicit, non-conscious processing of these subjects. Hence, the sole use of protocols—be it at an individual or at an agglomerated group level—risks the resemblance of linguistic descriptions that play a rhetorical function in describing and illustrating phenomena; examples of these fancy rhetorical functions, which simply cannot be fully validated by protocol analysis alone, are provided above (SUSTAINs and the like). Due to their intrinsic method constraints, it thus seems reasonable to aim toward combining protocols with other measures, and hence to avoid the risk of using one single metaphor only. More on this issue will be elaborated below.

## The Practice Metaphor

Finally, the last metaphor we identify as being frequently used in visual cognition analysis in medicine is the practice metaphor. Practice generates semantic structures of information that shape and are shaped by sequentially unfolding activity in relevant manners for a domain of scrutiny such as laparoscopy (Koschmann et al., 2007). As a starting point, we present the following quote from Carsetti (2004, p. 307) that we found interesting enough to use to introduce our reflection on the practice metaphor "A percept is something that lives and becomes, it possesses a biological complexity which is not be explained simply in terms of the computations by a neural network classifying on the basis of very simple mechanisms". This quote has two interesting elements. First, it emphasizes the lived nature of visual cognition, or what Livingston (1986) referred to as the lived work of reading. We will present an empirical example in the next paragraph that elaborates on this notion. Livingstone, in a series of ethnographic field descriptions, highlighted the sociability of practices that constitute intersubjective thinking and acting As such, the author provided a look that differed from looks "behind the skull" (Garfinkel, 1967) or from "computations by a neural network" (Carsetti, ibid). The second interesting element in this quote is that it seems to align neuroscientific work with labels such as "simply" and "very simple". We lack authority and motivation to judge such a judgment about the simpleness of neuroscience as being itself simplistic. Nevertheless, it illustrates the position of this author that something that focuses only on neural activation is unable to account for the full complexity of visual cognition (interestingly, compare the quote of Sauseng & Klimesch, 2008, starting the activation metaphor section, where the complexity of neural network communication is emphasized). Certainly, it is a matter of definition what "complex" is or what shall be allowed—based on which methodological and epistemological considerations—to have "complexity". Making such (maybe tacit, maybe deliberate) assumptions explicit is one of the purposes of this contribution.

To further illustrate the methodological entailments of the practice metaphor, we describe a recent example elaborating on the lived work of mammography. Slack, Hartwood, Procter, and Rouncefield (2007) highlight how diagnosing a mammogram is reflexively contingent on artful practices, in which multiple readers interact and intersubjectively constitute breast biographies. Central in their analyses are practices. Goodwin (1994, 2000) indicated that seeing and interpreting what is seen are not exclusively cognitive processes located in the individual brain (cf. activation and detection metaphors); rather, seeing is a socially situated activity accomplished through the deployment of a range of historically matured discursive practices. These practices constitute visual cognition in Goodwin's terms, and they are negotiated around a common object of disciplined perception (Lindwall & Lymer, 2008; Stevens & Hall, 1998), in the study of Slack and colleagues (2007): pictorial representations of the breast produced by an X-ray apparatus. Slack et al. identified practices such as arranging mammograms, manipulating, annotating, gesturing, and pointing that contribute to the lived work of doing a radiologic diagnosis. These representative practices (Greeno, 2006) unfold within an activity system, in many cases temporally over the course of minutes, but their sociogenesis stretches over the course of decades (such as the material resources used, i.e., pictures produced by X-ray technology). Hence, analysis of visual cognition using the practice metaphor adopts a different analytic time span than does, for example, analysis using the activation metaphor; and it adopts a sociotechnical system as the unit of analysis that explicitly accounts for the role of technology in medical image diagnosis (Burri & Dumit, 2008; Gegenfurtner, 2009). It is essentially the focus on embodied talk-in-interaction—talk between people and between humans and non-human objects (Gibson, 1979; Suchman, 2007)—that makes the practice metaphor a useful tool to analyze the fourth perspective; namely, what constitutes visual cognition, what medical students have to appropriate to move toward diagnostic accuracy, and how medical educators act to help develop this higher level of accuracy (see, e.g., Koschmann et al., 2007; Slack et al, 2007).

## Implications and Conclusions

This paper has presented and discussed a mapping of visual cognition analysis in medical education centered on four metaphors: activation, detection, inference, and practice. Reflection on these metaphors can give significant answers in three areas, namely: epistemology, methodology, and curriculum design. First, it provides an epistemological answer on the ongoing debate regarding the vision-cognition divide. Are perceptual and cognitive processes independent of each other? Or is the separation between vision and cognition artificial, a mere analytical distinction? Specifically, many studies in medical research adhering to the detection metaphor include no additional measures that are more prevalent in the inference metaphor, and vice versa. This neglect to combine measures for less fallibly accounting for visual cognition indicates the (maybe tacit) assumption in the medical literature that vision and cognition are independent. An assumption of this kind results from a way of thinking informed by an information processing view of cognition. This view stands in contrast to an ecological view on visual perception and cognition (Gibson, 1979), in which seeing, thinking, and acting are mutually interactive and nondualistic. It is questionable which approach is better: the information processing approach or the ecological approach. We argue here, as we have at the beginning of this paper, that it is not a question of truth or value judgment. Different methods lead to different answers based on different indicators. It is a reflection on these indicators, and more generally on the methodological entailments behind seemingly different metaphors, that can help raise awareness of each metaphor's (epistemological and pragmatic) potentiality and contingency, and that can thus advance our research practice in medical education and the learning sciences.

The second answer that is informed by the reflection on different metaphors on visual cognition concerns the dangers of choosing just one. Sfard (1998) indicated that a combination of learning metaphors yields to more robust findings than does a non-combination. This is in line with research in times that have been labeled the *decade of synergy* (Bransford et al., 2006). Yet, combining methods is neither trivial nor simple. As a first step, we briefly list three initial approaches for method synergy. First, eye-tracking and verbal reports can be combined to trace the interaction of perceptual and conceptual processing of visual information. Second, eye-tracking and EEG/fMRI recording can be combined to assess the first seconds of scanning an image—a period that is crucial, especially for higher levels of expertise. Third, video-recording and eye-tracking can be combined to measure the relationship between (1) mutual gaze patterns and (2) collaboration patterns as indicators of distributed visual cognition (Sangin et al, 2008; Stahl, 2006). In sum, synergies between metaphors need to be explored to avoid the dangers associated with choosing just one metaphor.

Finally, the third answer relates to curriculum design. It is thanks to the variety of metaphors available that we can illuminate ways to remake curricula in medical education. Specifically, the activation metaphor and its related methodologies inform the high speed with which visual information is processed, in the range of milliseconds; this cannot be captured with other methods. It also indicates the automaticity of information processing. Together with eye-tracking research, these findings indicate the role of implicit learning. In addition, the practice metaphor highlights the role of hands-on activities in interaction with the technological tool that is not merely used as a stimulus, but rather as a mediating object in the learning process. Learning to handle the tool is considered as important for medical education as is learning to discriminate features in the image; hence, curricula in medical schools can aim (or maybe have aimed) to also develop epistemic practices and professional identity. In summary medical curricula aimed at fostering implicit learning, identity formation, and participation can combine classroom-based direct instruction on disease etiology, learning in the lab with authentic tools, and work-based learning in real-life clinical workplaces, in ways that best support the development of "a good eye" in medial image diagnosis.

The ICLS 2010 conference theme invited an exploration of the ways disciplinary perspectives can inform the study of learning in educational or workplace settings. This theme—elaborated in the call for proposals—suggests an awareness of the multispectral image of learning caused by different, discipline-specific thinking traditions and their associated research custom. This paper has attempted to take up this theme; it did so by bringing together tradition and custom that have evolved over the years around a common topic: the analysis of visual cognition in medical image diagnosis. It can be argued that this disciplinary view, a medical education view, has implications for the study of learning that go beyond medicine. Specifically, in mathematics education, similar research practices can be identified that have as underlying inspiration the activation metaphor (e.g., when children focus spontaneously on number aspects during encoding of photos; Hannula, Grabner, & Lehtinen, 2009); the detection metaphor (e.g, when elementary students solve addition and subtraction word problems; De Corte, Verschaffel, & Pauwels, 1990); the inference metaphor (when schema-based instruction facilitates algebraic thinking for students with learning disabilities; Xin, 2008); or the practice metaphor (when participating in disciplined mathematical practices around coordinate systems; Stevens & Hall, 1998). These snapshots from mathematics education illustrate that the four metaphors identified in medical education literature might also have traces in other domains; if a more in-depth analysis of mathematics education research corroborates this initial snapshot finding, then this would answer one of the questions posed in the ICLS 2010 call for proposals—that is, what seems to be constant in learning processes (more specifically: in visual learning processes) beyond intrinsic disciplinary variations can be captured by different metaphors that highlight the multispectral image of our field: learning research.

## References


Alexander, P. A., Schallert, D. L., & Reynolds, R. E. (2009). What is learning anyway? A topographical perspective reconsidered. *Educational Psychologist, 44*, 176-192.

Ansari, D., & Coch, D. (2006) Bridge over troubled waters: education and cognitive neuroscience. *Trends in Cognitive Sciences, 10*, 146-151.

Arnheim, R. (1969). *Visual thinking*. Berkeley: University of California Press.

Augustyniak, P., & Tadeusiewicz, R. (2006). Assessment of electrocardiogram visual interpretation strategy based on scanpath analysis. *Physiological Measurement, 27*, 597-608.

Boshuizen, H. P. A., & Schmidt, H. G. (2008). The development of clinical reasoning expertise; Implications for teaching. In J. Higgs, M. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (pp. 113-121). Oxford: Butterworth-Heinemann.

Bransford, J., Stevens, R., Schwartz, D., Meltzoff, A., Pea, R., Roschelle, J., et al. (2006). Learning theories and education. Toward a decade of synergy. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 209-244). Mahwah, NJ: Erlbaum.

Brazeau-Lamontagne, L., Charlin, B., Gagnon, R., Samson, L., & Van der Vleuten, C. (2004). Measurement of perception and interpretation skills during radiology training: Utility of the script concordance approach. *Medical Teacher*, *26*, 326-332.

Burri, R. V., & Dumit, J. (2008). Social studies of scientific imaging and visualization. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *Handbook of science and technology studies* (pp. 297-317). Cambridge, MA: MIT Press.

Carsetti, A. (2004). The embodied meaning. Self-organization and symbolic dynamics in visual cognition. In A. Carsetti (Ed.), *Seeing, thinking, and knowing. Meaning and self-organization in visual cognition and thought* (pp. 307-327). Dordrecht: Kluwer.

Crowley, R. S., Lecowski, E., Medvedeva, O., Tseytlin, E., Roh, E., & Jukic, D. (2006). Evaluation of an intelligent tutoring system in pathology: Effects of external representations on performance gains, metacognition, and acceptance. *Journal of the American Medical Informatics Association*, *14*, 182-190.

Dawes, T. J. W., Vowler, S. L., Allen, M. C., & Dixon, A.K. (2004). Training improves medical student performance in image interpretation. *British Journal of Radiology*, *77*, 775-776.

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology*, *82*, 359-365.

De Jong, T., Van Gog, T., Jenks, K., Manlove, S., Van Hell, J., Jolles, J., et al. (2009). *Explorations in learning and the brain. On the potential of cognitive neuroscience for educational science*. Berlin: Springer.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Cambridge University Press.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *79*, S70-S81.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

Gegenfurtner, A. (2009). What is seen on the screen? Exploring collaborative interpretation, representational tools, and disciplined perception in medicine. In A. Dimitracopoulou, C. O'Malley, D. Suthers, & P. Reimann (Eds.), *Computer supported collaborative learning practices: CSCL2009 community events proceedings* (pp. 71-72). Rhodes: International Society of the Learning Sciences.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*, 606-633.

Goodwin, C. (2000). Practices of seeing: Visual analysis: An ethnomethodological approach. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 157-182). London: Sage.

Greeno, J. A. (2006). Learning in activity. In R. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 79-96). Cambridge, MA: Cambridge University Press.

Haller, S., & Radue, E. W. (2005). What is different about a radiologist's brain? *Radiology*, *236*, 983-989.

Hannula, M. M., Grabner, R., & Lehtinen, E. (2009, March). *Neural correlates of spontaneous focusing on numerosity (SFON) in a 9-year-longitudinal study of children's mathematical skills*. Paper presented at the EARLI Advanced Study Colloquium on Cognitive Neuroscience meets mathematics education, Bruges, Belgium.

Issenberg, S. B., McGaghie, W. C., Gordon, D. L., Symes, S., Petrusa, E. R., Hart, I. R., et al. (2002). Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice. *Teaching and Learning in Medicine*, *14*, 223-228.

Kolodner, J. L. (1983). Towards an understanding of the role of experience in the evolution from novice to expert. *International Journal of Man-Machine Systems*, *19*, 497-518.

Koschmann, T., LeBaron, C., Goodwin, C., Zemel, A., & Dunnington, G. (2007). Formulating the triangle of doom. *Gesture*, *7*, 97-118.

Koschmann, T., & Zemel, A. (2009). Optical pulsars and black arrows: Discoveries as occasioned productions. *Journal of the Learning Sciences*, 18, 200-246.

Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., Bhattacharyya, A. K., Scott, K. M., et al. (2006). Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Human Pathology*, *37*, 1543-1556.

Kundel, H. L. (2007). History of research in medical image perception. *Journal of the American College of Radiology*, *3*, 402-408.

Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition, and decision-making in pulmonary nodule detection. *Investigative Radiology*, *13*, 175-181.

Kundel, H. L., Nodine, C. F., Conant, E. F., Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, *242*, 396-402.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Erlbaum.

Lindwall, O., & Lymer, G. (2008). The dark matter of lab work: Illuminating the negotiation of disciplined perception in mechanics. *Journal of the Learning Sciences*, *17*, 180-224.

Livingston, E. (1986). *The ethnomethodological foundations of mathematics*. London: Kegan Paul.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309-332.

Mello-Thoms, C. (2003). Perception of breast cancer: Eye-position analysis of mammogram interpretation. *Academic Radiology*, *10*, 4-12.

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, *325*, 284-288.

Morita, J., Miwa, K., Kitasaka, T., Mori, K., Suenaga, Y., Iwano, S., et al. (2008). Interactions of perceptual and conceptual processing: Expertise in medical image diagnosing. *International Journal of Human-Computer Studies*, *66*, 370-390.

Norman, G. R., Coblentz, C. L., Brooks, L. R., & Babcook, C. J. (1992). Expertise in visual diagnosis: A review of the literature. *Academic Medicine*, *67*, S78-S83.

Paavola, S., Lipponen, L., & Hakkarainen, K. (2002). Epistemological foundations for CSCL: A comparison of three models of innovative knowledge communities. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community. CSCL 2002 Proceedings* (pp. 24-32). Hillsdale, NJ: Erlbaum.

Patel, V. L., Arocha, J. F., & Zhang, J. (2005). Thinking and reasoning in medicine. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 727-750). Cambridge, MA: Cambridge University Press.

Patel, V. L., & Groen, G. J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science*, *10*, 91-116.

Sangin, M., Molinari, G., Nüssli, M., & Dillenbourg, P. (2008). How learners use awareness cues about their peer knowledge? Insights from synchronized eye-tracking data. *International Conference of the Learning Sciences Proceedings.*

Sauseng, P., & Klimesch, W. (2008). What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience and Biobehavioral Reviews*, *32*, 1001-1013.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *The mechanisms of thought processes*. London: Stationery.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, *27*, 4-13.

Simpson, S. A., & Gilhooly, K. J. (1997). Diagnostic thinking processes: Evidence from a constructive interaction study of electrocardiogram (ECG) interpretation. *Applied Cognitive Psychology*, *11*, 543-554.

Slack, R., Hartswood, M., Procter, R., & Rouncefield, M. (2007). Cultures of reading: On professional vision and the lived work of mammography. In S. Hester & D. Francis (Eds.), *Orders of ordinary action. Respecifying sociological knowledge* (pp. 175-193). Aldershot: Ashgate.

Smith, P. L., & Ratcliff, R. (2009). An integrative theory of attention and decision-making in visual signal detection. *Psychological Review*, *116*, 283-317.

Stahl, G. (2006). Group *cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Stevens, R., & Hall, R. (1998). Disciplined perception: Learning to see in technoscience. In M. Lampert & M. L. Bunk (Eds.), *Talking mathematics in school: Studies of teaching and learning* (pp. 107-149). New York: Cambridge University Press.

Suchman, L. (2007). *Human-machine reconfigurations. Plans and situated actions* (2nd ed.) Cambridge, MA: Cambridge University Press.

Xie, Y. P. (2008). The effect of schema-based instruction in solving mathematics word problems: An emphasis on prealgrebraic conceptualization of multiplicative relations. *Journal for Research in Mathematics Education*, 39, 526-551.

## Acknowledgments

This work was supported by Grant No. 128766 of the Academy of Finland; it is part of an interdisciplinary research project at the confluence of medical sciences and learning sciences investigating “Technologies for seeing and technologies for knowing: Learning and medical imaging (LearnMedImage)”. In this project, research seldom is an individual endeavor, but more often is a collaborative achievement. Thus, we would like to thank Roger Säljö, Erno Lehtinen, Hans Rystedt, Laura Helle, Markus Nivala, and Minna M. Hannula for developing with us a good eye for metaphors with which to analyze visual medical cognition.

# Delinquent or criminal? Fostering conceptual understanding of technical terms in computer-mediated collaborative learning

Elisabeth Paus, University of Muenster, Fliednerstraße 21, 48149 Muenster, e.paus@uni-muenster.de
Gisela M. Gerhards, University of Frankfurt, Senckenberganlage 15, 60054 Frankfurt, giselagerhards@web.de
Regina Jucks, University of Muenster, Fliednerstraße 21, 48149 Muenster, jucks@uni-muenster.de

**Abstract:** To ensure a complex understanding of the topics addressed in collaborative learning scenarios, learners have to develop a conceptual understanding of the relevant specialist vocabulary. To this end, they have to exchange and discuss their individual knowledge, which is usually gained through individual exploration of different learning materials. Based on the psycholinguistic concept of lexical alignment, the following study analyzes whether providing learners with different technical terms in their respective learning materials has a positive impact on information exchange, term understanding, and learning. Results showed that dyads whose materials contained different terminology exchanged more information and performed better in a subsequent knowledge test than did dyads whose materials contained the same terminology. The inclusion of illustrations containing different terminology enhanced these effects. Implications for the selection and design of learning materials as well as perspectives for further research are discussed.

## Introduction

In general, language serves as a main tool by which thoughts, needs, attitudes, etc., are expressed. The main challenge in all verbal and written interactions is to achieve mutual understanding. This holds both for everyday situations, such as referring to a person or a thing when telling a story, and for more formal situations, such as referring to the conceptual knowledge that is usually encoded in specialist vocabulary in peer-to-peer learning situations (Bromme, Rambow, & Nückles, 2001). Focusing on the second scenario, let us imagine a typical situation in the life of university students: Peter writes Linda a message to coordinate the content of a presentation they are preparing for a course dealing with Freudian theory. He writes: "… and suppression is another important defense mechanism we should mention!" Linda answers: "I agree. I'll add suppression to the defense mechanisms." They probably both know the terms "suppression" and "defense mechanism" because they have read various texts to prepare for this exchange and the subsequent presentation. From the perspective of classic communication psychology, this discourse sequence clearly seems to be successful. Peter introduces two central terms and Linda signals understanding by using the terms without asking for clarification of their meaning or showing any other signs of mis- or non-understanding.

Following Clark and colleagues, the process by which interlocutors try to establish that messages have been received and understood is called grounding (Clark & Brennan, 1991). Grounding is an ongoing process during communication that serves to determine, to update, and to expand the so-called common ground (e.g., Clark, 1992; Clark & Marshall, 1981), that is, the assumed shared knowledge that provides the basis for communication. The common ground is inferred on the basis of different types of evidence; Clark and Marshall (1981) described three heuristics that people use to estimate what an interlocutor knows. Because in our example Peter and Linda communicate in written form and thus cannot rely on verbal or visual signs of mutual understanding (Monk, 2003), their assumptions about their shared knowledge are based mainly on word use (*linguistic co-presence heuristic*). In ongoing communication, newly introduced terms are usually repeated; they "become part of a 'discourse record' that is a subset of interlocutors' common ground" (Barr & Keysar, 2006, p. 6). The use of the same vocabulary is called lexical alignment (Pickering & Garrod, 2005) and takes place in both spoken (Barr & Keysar, 2002) and written communication (Bromme, Jucks, & Wagner, 2005). According to Clark and Marshall (1981), it is reasonable and efficient to stick to the lexical decisions made by the interlocutor in order to express one's own intended meaning. Interlocutors thus "minimize their collaborative effort—the work that both do from the initiation of each contribution to its mutual acceptance" (*principle of least collaborative effort*, Clark & Brennan, 1991, p. 135). Lexical alignment has also been demonstrated for technical terms in technical communication (Bromme et al., 2005; Jucks, Becker, & Bromme, 2008) such as that engaged in by Peter and Linda in our example. However, in the context of technical terms – as opposed to everyday terms – the difficulty arises that a term's connotations and complexity have to be transmitted within the choice of words and that an important part of the exchange concerns the underlying concepts. For example, whereas interlocutors in everyday communication may have different concepts of the meaning underlying the term "man," this is not relevant for successful communication ("The man is standing on the street."). In contrast, when interlocutors refer to the meanings underlying technical terms, as in our example, lexical alignment can conceal different meanings of a term that are relevant for discourse. Thus, where the deep understanding of the meaning underlying a term is concerned (and not simply the reference to an object in the world), lexical alignment may obscure the fact that the interlocutors have different conceptualizations of a term.

It has already been demonstrated that interlocutors' comprehension of the meaning of a technical term may be differently elaborated (Bromme et al., 2001). In particular, research on expert – layperson communication has shown that the underlying meaning of a technical term and the interlocutors' mental representation may differ (Jucks & Bromme, 2007). Whereas experts are usually aware of the complex knowledge encapsulated by technical terms (Schmidt & Boshuizen, 1992), laypersons' understanding of their meaning is often rather vague or even erroneous (Gittelman, Mahabee-Gittens, & Gonzalez-del-Ray, 2004). When interlocutors use the same words, differences in their knowledge of the underlying concepts are very difficult to detect. In general, it is difficult for people to distinguish between their own and their partner's knowledge of a particular subject, as has been demonstrated experimentally by Keysar (1994). In the same vein, Nickerson (1999) has shown that people tend to impute their own knowledge to others and only gradually become aware of others' knowledge states. Gauging others' conceptual knowledge about technical terms poses an additional difficulty: Because the underlying meaning is usually rather complex and the context of definition may vary (Paus & Jucks, 2009), it is very difficult for interlocutors to define an explicit criterion for knowing "enough" about a term.

Let us now return to our example: Peter introduces the words "suppression" and "defence mechanism." Linda absorbs these words without showing any signs of mis- or non-understanding – lexical alignment occurs. However, the assumption of shared knowledge, or common ground, may be mistaken, because Peter and Linda do not necessarily share the same knowledge of the underlying conceptual meanings. Consequently, using the same words can lead to particularly ineffective communication: Interlocutors communicate based on an illusion of common ground that prevents them from exchanging their knowledge explicitly and thereby developing a differentiated understanding of the concepts underlying the specialist terms. Lexical alignment of technical terms can thus be expected have certain costs in the area of learning and understanding. Indeed, Paus and Jucks (2008) have shown that the use of the same lexical encoding may—tacitly—be taken as an indicator of common ground in computer-mediated peer-to-peer learning settings. In an experimental study, university students working on a cooperative learning task gained more individual knowledge and showed more elaborated discourse patterns when the two participants in a learner dyad were provided with different terminology describing the same technical concept. These findings can be interpreted as follows: If linguistic co-presence is interrupted because the learning partners lack a joint vocabulary, their attention is explicitly drawn to their own and their partner's understanding of the terms. Not having the same words available allows them to assume less overlap in knowledge (Wu & Keysar, 2008). To increase their shared knowledge, they discuss the meaning of the terms. Individual knowledge about the underlying conceptual meaning is therefore exchanged more explicitly, thus fostering the development of a differentiated understanding of the meaning of technical terms as well as a better understanding of the topic in general. To conclude, learners' attention has to be focused on their understanding of technical terms.

Research has already shown that scientific illustrations in text materials attract readers' attention (Jucks, Bromme, & Runde, 2007) and "… exert a specific impact on the form and content of replies" (pp. 213-214). Providing learners with content-specific visualization tools is also known to have a positive effect on both communication and problem solving (Fischer, Bruhn, Gräsel, & Mandl, 2002). Nevertheless, the use of a term unfamiliar to the learning partner may initially cause confusion (Wu & Keysar, 2008). Although this confusion should prompt the more explicit elaboration of the meaning of specialist terms, interlocutors may not be aware of the resulting knowledge gain and instead feel a higher level of uncertainty.

Based on these considerations, our hypotheses for the present research were as follows: (1) We expected learning partners who were provided with different terms encoding the same underlying concepts to exchange more information and to gain a better understanding of the terms and the learning content than learning partners who were provided with the same terms. (2) Additionally, because learners' attention is attracted to the information provided in illustrations, we expected the predicted effects to be increased by the inclusion of illustrations in which the main concepts were also encoded by different technical terms. (3) Finally, we examined the effect of different terminology on learners' feelings of uncertainty. We expected learning partners provided with different technical terms to feel more uncertainty about their own knowledge than learning partners provided with the same terms.

# Method

## Participants and Design

Participants were 72 undergraduate psychology students (47 female, 25 male) at the University of Frankfurt. Mean age was 24.37 years (*SD* = 4.63). All participants were German native speakers. They were assigned to 36 learning dyads and randomly assigned to three conditions: texts with the same terminology, texts with different terminology, texts with an illustration and different terminology. Participation was voluntary and rewarded with 7 euro. Learning outcomes were assessed for each participant individually (sample size: 72

participants, with 24 in each condition); the communication process was analyzed within dyads (sample size: 36 dyads, with 12 in each condition).

## Procedure

Two students participated in each trial. They were tested individually in separate rooms and were assigned gender-neutral pseudo-email addresses to prevent them from making any assumptions about each other. Note that we decided on email rather than chat as the communication medium due to practical aspects of data collection. Each room was equipped with the necessary technology and an experimenter was present throughout. Access to external Internet sites was blocked to prevent participants from using information sources other than the texts provided. First, it was ensured that the participants' computer skills met the demands of the experiment. The partners in each dyad were then given different texts to read and understand with no time limit. They were then set a collaborative task, to be discussed and completed with their partner at their own pace via e-mail. To ensure that the learning scenario was as realistic as possible, no further instructions were given on the procedure and/or the collaboration process (e.g., length, structure, etc.). After the collaborative task, the learning materials were removed to ensure independent knowledge generation in the knowledge tests. Participants provided with an illustration were asked to rate the perceived importance of the illustration for solving the collaborative task. Moreover, all participants were asked to rate their feeling of uncertainty with regard to their knowledge of the topic covered. Afterwards, they were administered a knowledge posttest comprising a cloze task (Taylor, 1957) and a multiple-choice test. Finally, demographic data including the frequency of computer usage were obtained.

## Materials

### Learning texts

The learning texts were original extracts from two child psychiatry textbooks (Herpertz-Dahlmann, 2005; Schmeck, 2004). Both extracts dealt with the development of a social behavior disorder; each was around 500 words long. Except for some abridgements to ensure that the two texts had the same length and format, no modifications were made. Although both texts addressed the same topic, there were differences in the description of theoretical models and in the detail of the information presented. Both texts covered all information assessed in the knowledge posttest.

### Manipulation of the learning material

We first drew up a list of terminology specific to the topics covered. Fifteen key concepts were identified. Some of the terminology used to designate the same underlying concept in the original texts differed (e.g., prevalence vs. frequency); some of it was the same (e.g., monozygotic). In cases where the terminology was the same in both texts, we looked for synonyms in a dictionary of foreign words (Wermke, Klosa, Kunkel-Razum, & Scholze-Stubenrecht, 2001). Each term was matched with a synonym that had the same meaning in the present context. Two different versions of the text material were then constructed by manipulating the usage of the 15 key concepts. In the different terminology condition, the 15 key concepts were used in one encoding (e.g., prevalence) in one text and in the other encoding (frequency) in the other text. In the control condition, the same terminology (prevalence) was used in both texts. When matching the words, we sought to maintain a balance across the texts in the use of terms stemming from Latin or Greek and terms with German roots. The illustrations were both taken from the respective textbooks. They contained 5 vs. 6 of the 15 manipulated terms as well as further terms relating to the development of a social disorder. The terms in the illustration were adapted to the terminology used in the respective version of the text.

### Collaborative task

We set participants two questions to be worked on collaboratively. The questions were designed to support the integration of factual knowledge on the topic at hand as well as the production of inferences during the learning process. To answer the first question, they had to refer to the case study of a 13-year-old boy with behavioral problems. They were asked to list all factors that might have influenced his behavior. The second question concerned the prevalence of social behavioral disorders in boys and girls. Learning partners had to exchange the information contained in their individual texts in order to succeed in the collaborative task.

## Dependent Measures

Three sets of outcome measures were considered: information exchange, individual knowledge gain, and subjective measures.

### Information exchange

Three measures were used to assess information exchange within the dyads:

*Communication.* To assess information pooling during communication, we determined the number of questions asked and the number of explanations given within each dyad. We restricted the assessment to questions and explanations relating to the relevant terminology. Participants' questions and explanations were transcribed and analyzed (by a second rater) according to predefined criteria. Interrater reliability for the numbers of questions asked ($K = 1.0$, $p < .001$) and the number of explanations given ($K = 1.0$, $p < .001$) was satisfactory.

*Knowledge exchange items.* The multiple-choice test included questions tapping knowledge of contents that were included in only one version of the text (six questions for each text). These items were intended to measure information pooling between the learning partners, regardless of the usage of technical terms. Learning partners could answer these items correctly only if they had exchanged their individual knowledge. Additionally, for the two conditions in which learners were provided with different terms, we counted the number of technical terms from the partner's learning material that were used in a cloze procedure to determine how many of these terms had been absorbed from the partner.

### Individual learning outcomes

Individual knowledge gain was measured by the following indicators:

*Multiple choice test:* General understanding of the contents covered was assessed by a multiple-choice test with 28 tasks, each with four response alternatives. All distractors were chosen such that an uninformed participant perceived them as being correct with the same probability (Bortz & Döring, 1995). One point was given for each correct answer, and one point was subtracted for each incorrect answer. Hence, the maximum number of attainable points was 28.

*Cloze procedure.* We developed a cloze test requiring participants to fill in words that had been removed from a text to determine whether they could reproduce the topic-relevant terminology in a continuous text, thus demonstrating an understanding of the underlying concepts (Taylor, 1957). A rational deletion procedure was used to identify the 10 words that were omitted; only content words (subjects, objects, adjectives, and verbs) were removed (Kobayashi, 2002). Eight of the omitted words were drawn from the list of manipulated terminology. We were thus able to test the participants' understanding of the terms. The remaining two items tested the main topics covered in the two texts. Participants were given one point for each gap filled correctly. Thus, the maximum number of attainable points was 10.

### Subjective ratings

*Relevance of illustration.* Participants were asked to rate the perceived relevance of the illustration on a 5-point scale ranging from *not true* to *true*.

*Feeling of uncertainty*. Likewise, participants were asked to rate their perceived uncertainty on a 5-point scale from *not true* to *true* to determine whether the provision of different terminology influenced their feeling of uncertainty.

## Results

In this section, we report the effects on the three sets of outcome measures. Unless otherwise indicated, analyses were performed using SPSS 17. Participants in the three conditions did not differ in frequency of computer usage, $F(2,69) = 0.92$, $p = .402$, frequency of internet usage, $F(2,69) = 0.14$, $p = .873$, or subjective self-ratings, $F(2,69) = 0.64$, $p. = .53$.

## Information Exchange

*Communication.* The length of communications was $M = 587.36$ words ($SD = 215.37$); there was no effect of condition, $F(2,33) = 3.10$, $p > 0.05$, *ns*. MANOVAs with the number of questions/explanations relating to the manipulated vocabulary as dependent variables revealed a main effect of condition, $F(4,66) = 2.79$, $p < 0.05$, $\eta_p^2 = .15$. However, a univariate analysis revealed no differences in the number of questions across the three conditions, $F(2,33) = .88$, $p > 0.05$, *ns*. In contrast, a univariate analysis of the number of explanations revealed a main effect of condition, $F(2,33) = 6.67$, $p < 0.05$, $\eta_p^2 = .15$, with participants in the different-terminology with illustration condition giving more explanations ($M = 5.92$, $SD = 3.0$) than participants in the different-terminology without illustration condition ($M = 3.92$, $SD = 3.2$) or participants in the same-terminology condition ($M = 1.67$, $SD = 2.27$). Most explanations related to the terms "prevalence," "frequency," "social pressures," "age group," "adolescence," "psycho-social risk factors," and "genetic."

*Knowledge exchange items.* An ANOVA revealed an effect of condition on the number of correctly solved multiple-choice items designed to measure information exchange, $F(2,69) = 11.17$, $p < 0.001$, $\eta_p^2 = .25$. A post hoc test (S-N-K) showed that the same-terminology group differed from the different-terminology group with illustration ($p < 0.05$) and from the different-terminology group without illustration ($p < 0.05$), whereas no difference was found between these two groups ($p > 0.05$, *ns*).

In the cloze task, participants in the different-terminology condition used *M* = 0.75 (*SD* = 0.98) terms from their partner's learning material to fill the gaps and *M* = 3.25 (*SD* = 1.90) terms from their own material. However, the difference between groups in term usage was not significant, $t(46) = -.29$, $p > 0.05$, *ns*.

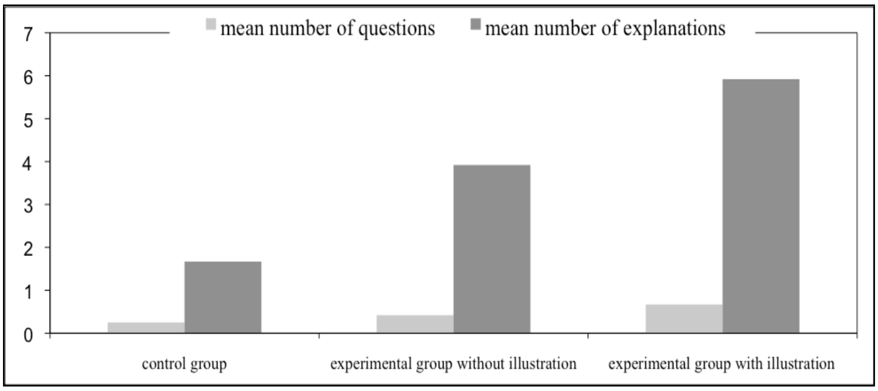



Figure 1. Mean number of questions asked and explanations given in the communication process

## Individual Learning Outcomes

To examine the effects of condition on students' learning performance, we conducted a MANOVA on the number of points attained in the multiple-choice test and in the cloze test. Pillai's trace showed an effect of condition, $F(4,138) = 6.65$, $p < 0.001$, $\eta_p^2 = .16$.

*Multiple-choice test.* On average, participants scored 14.46 (*SD* = 7.61) of the available 28 points. Univariate analysis revealed a main effect of condition on points attained in the multiple-choice test, $F(2,69) = 12.8$, $p < 0.001$, $\eta_p^2 = .27$. A post hoc test (S-N-K) revealed that the same-terminology group differed from both different-terminology groups ($p < 0.05$). Moreover, the two different-terminology groups differed from each other ($p < 0.05$).

*Cloze test.* On average, participants correctly filled in 5.46 (*SD* = 2.53) of the 10 gaps. An ANOVA showed a main effect of condition, $F(2,69) = 10.53$, $p < 0.001$, $\eta_p^2 = .23$. A post hoc test (S-N-K) revealed that the same-terminology group differed from both different-terminology conditions ($p < 0.05$). However, no difference was found between the different-terminology conditions with and without illustration ($p > 0.05$, *ns*).

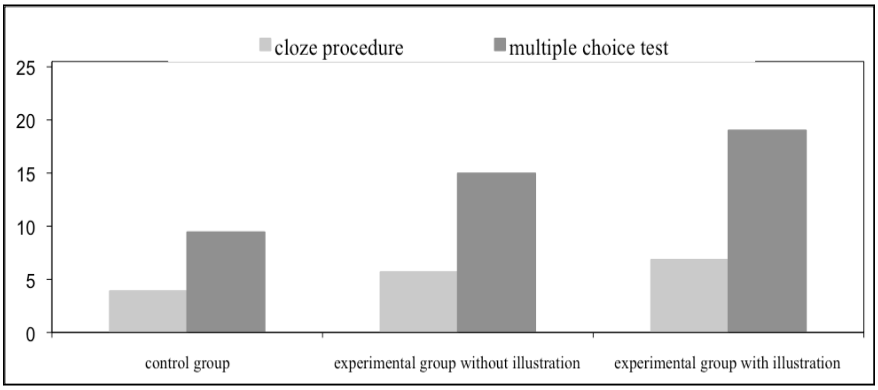



Figure 2. Mean scores in the cloze test and multiple-choice test

## Subjective Ratings

*Relevance of illustration.* On average, the illustration was rated as being rather helpful (*M* = 3.46, *SD* = 1.41). To test how this perception related to objective learning outcomes, we correlated the subjective rating of the illustration's relevance for learning with the objective learning outcome in the multiple-choice test. The correlation was not significant, $K = .251$, $p > 0.05$, *ns*.

*Feeling of uncertainty*. Participants reported little uncertainty about their knowledge (*M* = 1.71, *SD* = 0.86). An ANOVA revealed no difference between the three groups, $F(2, 69) = 0.39$, $p > 0.05$, *ns.*

## Discussion

These results emphasize the pivotal role played by technical terms in the development of a conceptual understanding of learning contents. We hypothesized that presenting learners with materials containing different lexical encodings of the same concept would initiate more reflection on their own and their partner's knowledge. We therefore expected learners to realize that their knowledge differed from that of their partner, which would prompt them to exchange knowledge and to elaborate the meaning of the technical terms in more depth. In the following, we discuss our findings with regard to information exchange, individual learning outcomes, and additional subjective self-ratings.

### Information Exchange

In accordance with our hypothesis, learners in the different-terminology condition with an illustration made more explanatory statements about this terminology than did learners in the same-terminology condition. Taking into account that our experimental manipulation was limited to the lexical encoding of 15 key concepts, we interpret these results as evidence for the role of word choices in learning situations. The findings indicate that exposure to different terminology prompts learners to communicate in a more productive way, exchanging information about the underlying meaning of the terms. We attribute this difference to the use of different terms in their communication on the collaborative task making participants being more aware of potential problems of understanding. In particular, the results highlight the specific role of illustrations for information exchange in the given learning setting. As hypothesized, illustrations strongly focused the learners' attention on their own learning material, making them more aware of differences in the terminology used. To establish a common ground, they provided an explanation for the terms they used. In contrast, the explicit exchange of information did not seem necessary for dyads in the same-terminology condition, who assumed their partners to have similar understanding of the term as themselves (*knowledge effect*; Bromme et al., 2001).

Moreover, learners who exchanged more information were more familiar with the contents of their partner's learning material: Participants in the different-terminology conditions were better able to solve the multiple-choice items designed to measure information exchange. However, the same did not apply to recall of technical terms: In the cloze task, participants clearly reverted to the technical terms used in their own material. Hence, whereas increased information exchange led to more knowledge about non-shared contents, learners stuck to their own material on the lexical level.

### Individual Learning Outcomes

We examined learners' individual learning outcomes in terms of their conceptual understanding of the topic and their understanding of the technical terms used. Performance on the multiple-choice test differed across all three groups, with participants in the different-terminology with illustration condition achieving the highest scores. These findings support our hypothesis that elaboration of learning content can be improved by placing different vocabulary in cooperative learning texts and that this effect can be further enhanced by including an illustration. Besides focusing learners' attention on the technical terms used, illustrations also help them to better understand the relations between the concepts (Larkin & Simon, 1987). In the cloze task, participants in the different-terminology conditions outperformed participants in the same-terminology condition. However, provision of an illustration had no additional influence here. In conclusion, it seems that inclusion of an illustration using different terminology to encode the same concepts has a positive effect on the development of a differentiated understanding of learning contents, but does not particularly affect term understanding.

Participants also reported that the illustration helped them to understand the learning contents. However, this perception was not systematically related to objective learning outcomes.

### Limitations and Future Perspectives

In contrast to our hypothesis, participants' subjective ratings of uncertainty did not differ across conditions: Participants in all conditions reported little uncertainty about their knowledge. It is likely that the meaning of most of the manipulated terms could simply be deduced from the context. This interpretation is supported by the finding that the number of questions asked to clarify the underlying meaning of terms did not differ across conditions either. Participants only explicitly queried the meaning of very specific terms that are not used in everyday language. It seems that communication partners exchanged knowledge not because they were unable to understand each other, but because the different terminology used in the learning materials signaled different knowledge and thus led to perspective taking. Participants realized that their knowledge was not necessarily the same and thus started to explain their individual understanding of the term. It seems likely that this spontaneous provision of explanations directly reduced uncertainty and, in turn, the need to request more details. It remains an open question whether the lack of questions formulated can be attributed to characteristics of the terms or to the fact that the meaning was already clarified. Paus & Jucks (2009) have shown that learners' perceived understanding is affected by the origin of terms (e.g., Greek, Latin, Germanic).

In Germany, terms with Germanic roots are more easily accessible, perceived to be more familiar, and better understood. To address this question, the characteristics of the terms used in collaborative learning situations could be varied more systematically in future research. Individual factors, such as the degree of expertise in the field (see Gittelman et al., 2004; Schmidt & Boshuizen, 1992) or foreign language skills, may also affect perceived understanding of terms. Additionally, the correspondence between subjective assessments of the cooperation process and the perceived difficulty of the learning materials warrants analysis. It is important to find out whether and how learners perceive themselves to learn in collaborative learning environments and to identify how much pressure they face in coming to a mutual understanding. Future research should thus assess the mental processes underlying the learners' behavior; for instance, in retrospective analyses. Last but not least, it remains unclear what role social expectancy plays in admitting a feeling of uncertainty about one's knowledge.

The experimental setting investigated in this study can be considered a typical learning situation at university. However, although learning partners often do not know each other, especially in distance universities, the question arises of how their level of acquaintance affects communication behavior. It can be assumed that learners are acting in social roles predefined by previous interaction situations and that motivational and social-psychological aspects of discourse behavior are strongly influenced by those roles. In the present study, for instance, participants made assumptions about their partner's gender based on his or her way of formulating messages and coordinating the interaction. Future research should clarify whether our findings can be transferred to other learning situations: What is the impact of group size? How does the medium influence communication behavior? Are there any differences in comparison to face-to-face learning scenarios? Characteristics of the learning partners should also be taken into account. Spiro, Coulson, Feltovich, and Anderson (1994) argued that humanities students can be expected to use heuristic procedures to solve problems and to debate different alternatives. These students can thus be able to recognize that two terms may exist to describe the same underlying concept. In addition, learners' age can be expected to impact the development of assumptions about knowledge. Several researchers have concluded that this development proceeds stepwise (King & Kitchener, 1994; Kuhn, 1991; Perry, 1970). The validity of the present results should thus be verified for other learners, such as school students.

In summary, our findings indicate that the terminology used in learning materials impacts learning behavior in discourse. Moreover, the notation of knowledge contents (text vs. illustration) influences the way learners elaborate information. These aspects should therefore be considered in the design and selection of learning materials.

## References


Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46*, 391-418.

Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (Vol. 2, pp. 901-938). Amsterdam: Elsevier.

Bortz, J., & Döring, N. (1995): *Forschungsmethoden und Evaluation* (2nd ed.). Berlin: Springer.

Bromme, R., Jucks, R., & Wagner, T. (2005). How to refer to "diabetes"? Language in online health advice. *Applied Cognitive Psychology, 19,* 569-586.

Bromme, R., Rambow, R., & Nueckles, M. (2001). Expertise and estimating what other people know: The influence of professional experience and type of knowledge. *Journal of Experimental Psychology Applied, 7*, 317-330.

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: The University of Chicago Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA Books.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.). *Elements of discourse understanding* (pp. 10-63). Cambridge, UK: Cambridge University Press.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12,* 213-232.

Gittelman, M. A., Mahabee-Gittens, E. M., & Gonzales-del-Rey, J. (2004). Common medical terms defined by parents: Are we speaking the same language? *Pediatric Emergency Care, 20*, 754-758.

Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal for Research in Science Teaching, 24*, 291-307. [nicht im Text]

Herpertz-Dahlmann, B., Hagenah, U., Vloet, T., & Holtkamp, K. (2005). Essstörungen in der Adoleszenz [Eating disorders in adolescence]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *54*(4) 249-267.

Jucks, R., Becker, B.-M., & Bromme, R (2008). Lexical entrainment in written discourse: Is experts' word use adapted to the addressee? *Discourse Processes*, *45*, 497-518.

Jucks, R., & Bromme, R. (2007). Choice of words in doctor–patient communication: An analysis of health-related Internet sites. *Health Communication*, *21*, 267-277.

Jucks, R., Bromme, R., & Runde, A. (2007). Explaining with non-shared illustrations: How they constrain explanations. *Learning and Instruction, 17,* 204-218.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective-taking in text. *Cognitive Psychology, 26(2*), 165-208.

King, P. M., & Kitchener, K.S. (1994). *Developing reflective judgement: Understanding and promoting intellectual growth and critical thinking in adolescents and adults.* San Francisco, CA: Jossey-Bass.

Kobayashi, M. (2002). Cloze tests revisited. Exploring item characteristics with special attention to scoring methods. *The Modern Language Journal, 86*, 571-586.

Kuhn, D. (1991). *The skills of argument*. Cambridge, UK: Cambridge University Press.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Monk, A. (2003). Common ground in electronically mediated communication: Clark's theory of language use. In J. M. Carroll (Ed.), *HCI models, theories, and frameworks: Toward a multidisciplinary science* (pp. 265-289). San Francisco: Morgan Kaufmann.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737-759.

Paus, E., & Jucks, R. (2008). Do we really mean the same? The relationship between word choices and computer-mediated cooperative learning. In P. A. Kirschner, F. Prins, V. Jonker, & G. Kanselaar (Eds.). *International Perspectives in the Learning Sciences: Cre8ing a learning world. Proceedings of the 8th International Conference for the Learning Sciences* (Vol. 2, pp. 172-180).

Paus, E., & Jucks, R. (2009). Words and meaning. How the lexical encoding of technical concepts contributes to their mental representation. In N. Taatgen, H. van Rjn, L. Schomaker, & J. Nerbornne (Eds.). *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 496-501Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme.* New York: Holt.

Pickering, M. J.,. & Garrod, S. (2005). Establishing and using routines during dialogue: Implications for psychology and linguistics. In A. Cutler (Ed.) *Twenty-first century psycholinguistics: Four Cornerstones* (pp. 85-101). London, UK: Erlbaum.

Schmeck, K. (2004). Störungen des Sozialverhaltens [Disorders of social behavior]. In C. Eggers, J. M. Fegert, & F. Resch (Eds.) *Lehrbuch der Kinder- und Jugendpsychiatrie* (pp. 849-873). Heidelberg, Germany: Springer Verlag.

Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In D. A. Evans & V. L. Patel (E.). *Advanced models of cognition for medical training and practice* (pp. 265-282). Berlin, Germany: Springer.

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1994). Cognitive flexibility theory: Complex knowledge acquisition in ill-structured domains. In R. B. Ruddell, M. R.Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 602-615). Newark, DE: International Reading Association.

Taylor, W. L. (1957). Cloze readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology, 41*, 19-26.

Wermke, M., Klosa, A., Kunkel-Razum, K., & Scholze-Stubenrecht, W. (2001). *Duden – Das Fremdwörterbuch* [Dictionary of foreign words]. Mannheim, Germany: Brockhaus.

Wu, S., & Keysar, B. (2008). The effect of information overlap on communication effectiveness. *Cognitive Science*, *31*, 169-181.

## Acknowledgments

The authors are thankful to Stefanie Diener for her support with data collection and analysis

# Leading to Win: The Influence of Leadership Styles on Team Performance during a Computer Game Training

Anna Siewiorek, Andreas Gegenfurtner, Centre for Learning Research, University of Turku, Finland
Email: amsiew@utu.fi, angege@utu.fi

**Abstract:** The purpose of this study is to examine how leadership styles will influence the team performance during a computer game training. In order to get a better understanding what leadership styles are going to be distinguished, researchers observed students' interactions while they played a strategic computer game. In the study, a group of Stanford University graduate students participated in the training with a 'real esate' computer game. The participants' task during the game was to manage an estate company in small teams. Students developed goals, discussed problems and tracked progress in order to win the game. The results showed that the teams had different leadership styles which affected their performance. The leadership styles that participants used and how they affected the teams' performance are described in the paper.

## Introduction

Increasingly, employers are demanding that college graduates have integrative skills (Stephen, Parente, & Brown, 2002) for an interdisciplinary understanding of the organization and business dynamics (Hartenian, Schellenger, & Frederickson, 2001). However, methods and tools commonly used for teaching this business dynamic in collages are insufficient to cope with the complexity of organizations and the unstable conditions of today's market (Baker & O'Neil, 2002; Day, 2001; Lehtinen, 2002). Especially, it is difficult to teach leadership—a central competence for coping with these kinds of conditions—using traditional classroom-based methods of teaching (e.g., lectures). One solution for that problem is to teach about leadership styles through strategic computer games. Thus, more studies need to be conducted about the ways of teaching leadership styles in practice. It would be especially interesting to study how leadership styles influence team performance during computer gaming training. Previous studies have demonstrated significant linkages among leadership behavior, team efficacy, and team performance (Pescosolido, 2001). However, there has been a lack of integration concerning the relationship between specific leader behaviors and team performance outcomes (Burke et al., 2006).

This paper describes serious gaming session that aimed to teach students about business dynamics and about leadership styles i.e., which leadership style should be used in a specific situation to manage similar situations in their future workplace. Past studies have shown that serious games can successfully foster student learning (Kafai, 2006), engagement (Kiili & Lainema, 2008), and motivation (Hense & Mandl, 2009). It is worth noting that there is an evidence to suggest that playing games allows the brain to 'work' more efficiently and thus, to take in more cognitive material than it would in a more traditional setting (Jenkins, 2002; Perry & Ballou, 1997). In addition, games can be considered as powerful learning environments because of the following: (a) they are based upon active-learning techniques, (b) they favor activation of prior knowledge given that players must use previously learned information in order to advance, (c) they provide immediate feedback enabling players to test hypotheses and learn from their actions, (d) they encompass opportunities for self-assessment through the mechanisms of scoring and reaching different levels, and (e) they provide opportunities for critical thinking and problem solving skills (Oblinger, 2004). Furthermore, few examples of using simulation games to develop leadership skills can be found in the military sector (O'Neil & Fisher, 2004). The military sector also uses simulation-based games in flight and combat training (O'Neil & Andrews, 2000). For this study, a strategic computer game, "Build-a-lot" (see Appendix), was used in which players met face-to-face in small teams to manage a virtual real estate company. Participants collaborated in teams in order to win the game, and group work helped them to share and develop alternative viewpoints. Furthermore, the use of teams to promote student learning in education has been proven to be fruitful (Michaelsen, Bauman Knight, & Fink, 2004). For example, in Light's study (2001) of student learning at Harvard, he found student learning teams to be highly effective. On the contrary, small group research has identified a number of factors that negatively affect team performance and member satisfaction. These include over-dependence on a dominant leader (Edmondson, Bohmer, & Pisano, 2001).

The computer game used for the study presented an enormous amount of complex operations in which a team leader needed to address the final decision to lead the team to win the game. This kind of computer game, involving many complex decisions, can connect the players to everyday life experiences. In addition, the time pressure in the game was crucial, and participants had to make decisions quickly in order to advance to the next level. The speed of the game was fast, and only quick decision-making could lead teams to win the game. This gaming environment was interactive, and participants interacted by solving the given tasks in the game.

Therefore, strategy and an appropriate leadership were needed in order to be successful. Furthermore, the training was competition-based and that also required implementation of leadership skills. The research questions were as follows: 1) what leadership styles (if any) could be distinguished during the computer game, and 2) how have leadership styles influenced team performance during the game?

## Leadership styles

Most definitions of leadership claim that it is a process of intentional influence by one person over others "to guide, structure, and facilitate activities and relationships in a group or organization" (Yukl, 1998, p. 3). In other words, leadership is defined in terms of problem-solving activities directed at the generation of solutions that advance team goal attainment (Zaccaro, Rittman, & Marks, 2001). Leadership is typically associated with more mystical, charismatic qualities such as the ability to influence, arouse, inspire, enthuse, and transform (Bass & Avolio, 1994; Goleman, 2000), qualities that exist outside of the ordinary and mundane features of daily life. Additionally, leadership is frequently theorized as the exercise of power, the setting of goals and objectives, the managing of cultures, and the mobilization of others to get work done (Alvesson & Sveningsson, 2003; Gemmill & Oakley, 1992). In addition, leadership has been defined in terms of individual personality traits, leader behaviors, responses to leader behaviors, interpersonal exchange relationships, interaction patterns, role relationships, follower perceptions, task goals, organizational culture, and nature of work processes (Mello, 2003; Rost, 1991; Yukl, 1989). Hence, leadership is an important component in business dynamics. Below, we specify four leadership styles in more details: transactional, transformational, heroic, and post-heroic leadership.

First, transactional leadership is characterized by the clear specification of what followers are expected to do (e.g., Bryman, 1996) and it is based on a rational exchange relationship between leader and subordinate (Bass, 1985; Howell & Costley, 2001). The leader articulates what behaviors are required and what will be rewarded, and provides feedback to the subordinates about their behaviors. The subordinate, in turn, complies with these behavior requirements if rewards are desired (Yun, Cox, Sims, & Salam, 2007). In short, transactional leadership can be defined as a cost-benefit exchange between leaders and their followers and it is comprised of three dimensions: contingent reward, active management by exception, and passive management by exception.

On the other hand, transformational leadership is characterized by concern with people. Transformational leaders lead by inspiring and stimulating followers; by creating highly absorbing and motivating visions; and by utilizing behaviors such as charisma and intellectual stimulation to induce performance of subordinates beyond expectations (Conger, Kanungo, & Menon, 2000; Kark, Shamir, & Chen, 2003; Yun et al., 2007). Bass (2000, p. 21) states that transformational leaders "move followers to go beyond their own self-interests for the good of their group, organization or community, country or society as a whole". Transformational leadership is comprised of the following dimensions: idealized influence, inspirational motivation, charismatic and intellectual stimulation, and individual consideration.

Third, heroic leadership is characterized by the following: feelings that leadership is based on superior knowledge and information (omnipotence); fearing failure (rightness); keeping up appearances at any cost including blaming others (face-saving); and viewing subordinates as inferior creatures in constant need of assistance and rescue (codependency) (Eicher, 2006). Two sub-classes could be distinguished in heroic leadership such as: autocratic and coercive leadership styles which have many similarities. The autocratic leader sequesters a high degree of control and makes decisions without consulting team members. The heroic-autocratic leader determines all policies, techniques, and activity steps, and dictates the particular work tasks and work companions of each member. The autocratic leader tends to be personal in her praise and criticism of the work of each member but remains aloof from active group participation (Choi, 2007). On the other hand, in coercive leadership style leader demands immediate compliance and drive to achieve, initiative, and self-control (Goleman, 2000).

Finally, post-heroic leadership takes place when the leader wants others to take responsibility and gain knowledge (empowerment), encourages innovation and participation even in ambiguous situations (risk-taking), seeks input and aims for consensus in decision-making (participation), and wants others to grow and learn even at the expense of becoming dispensable herself (development). Post-heroic leadership has thus become a concept used to describe a new conceptualization of leadership that refutes the top-down focus on the leader typical of most leadership literature and discourse (Bradford & Cohen, 1998; Eicher, 2006, Fletcher, 2004). This leadership style is well-suited to complex, changing, and inter-dependent environments. There are two sub-classes in post-heroic leadership: democratic and shared leadership. First, democratic leadership style emphasizes high group participation, discussion, and group decisions based on consensus encouraged by the leader. By giving workers a voice in decisions, democratic leaders build organizational flexibility and responsibility and help generate fresh ideas. In addition, because they have a say in setting their goals and the standards for evaluating success, people operating in a democratic system tend to be very realistic about what can and cannot be accomplished (Goleman, 2000). Second, shared leadership occurs when all members of a team are fully engaged in the leadership of the team and are not hesitant to influence and guide their fellow team members in an effort to maximize the potential of the team as a whole (Pearce, 2004). In summary, shared leadership represents teams whose members are empowered to share the tasks and responsibilities of leadership (Ensley, Pearson, & Pearce, 2003; Katzenbach, 1997).

## Methods

In order to get a better understanding of how leadership styles will affect team performance during computer game training, three researchers observed students' interactions when they were playing a strategic computer game called "Build-a-lot" (see Appendix). Researchers were taking detailed notes of all student dialogs and actions; each researcher had one team to observe and take notes of their interactions and dialogs. The goal of this observation was to determine what leadership styles (if any) students are going to use during the game and how this styles will affect team performance. All notes that researched made were transcribed verbatim. For the purpose of data analysis the detailed codes were developed which described each leadership style characteristics. The notes (i.e. each team dialogs and interactions) were coded using those leadership styles characteristics. Nine participants took part in the gaming session. They were divided into three teams, which competed against each other for one hour. Final scores of the game determined which team won. During the game, students made many decisions related to running a real estate company. Due to the game's limited time, the participants became focused during the gaming session and were thoroughly engaged and immersed in the game. The time pressure increased the speed of their decision-making. The gaming session took place at Stanford University, California, USA in the autumn semester of 2007.

### Participants

Nine participants took part in the study (two females and seven males). They were Stanford University graduate students from different academic backgrounds such as engineering and humanities. The participants in the teams did not know each other. There were three teams with three participants in each team. They did not consider themselves experienced users of computer games. The study was the first in which all of them took part in a gaming session. During the experiment, participants worked together in teams because only in teams could leaders be distinguished. Participants were unaware of the purpose of the study.

### Description of the game and setting

Participants played "Build-a-lot" (see Appendix) in teams. Players represented real estate moguls whose task was to take over the housing market as they build, upgrade, sell, and buy houses for huge profits. The purpose of the gaming session was to teach students business dynamics, how to make decisions quickly. Their goal was to win the game; the team with the most profit and the highest achieved level won the game. From time to time, a message from the Mayor (of the city in which they were building houses) appeared on the computer screen with hints or tasks about the game. If a team managed to get every operation correct, they passed the level in the game and advanced to the next level. If a team failed a level in the game, they had to repeat that level from the beginning. Every team had one notebook computer to use along with a computer mouse. The teams played simultaneously for one hour. During the experiment, there were three researchers to observe the teams; each researcher had one team to observe. Researchers introduced the game to participants at the beginning and told them that they could look for help and instructions in the tutorial section in the game, but they could not ask the observing researchers any questions. Team members could ask questions of other teams. The aim of the researchers was to observe students' interactions while playing the game (how they collaborated, how they made decisions, etc.) and to make detailed notes of all their dialogs and students interactions.

### Data Analysis

The main aim of the study was to research how leadership styles which participants applied during the gaming session will influence the teams' performance. The method of data analysis was discourse analysis (Silverman, 2006). After reading through the transcripts of notes from observations and students dialogs, the transcripts were coded in terms of transactional, transformational, heroic and post-heroic leadership styles. These developed categories of leadership styles (each had subcategories referring to the given leadership style characteristics) were used to analyze the data.

## Results

The conclusion drawn from the observations is that the teams had different leadership styles, which affected their performance. The following three sections describe each team's leadership style and how it influenced performance and the game results.

### Team 1.

The leadership style of Team 1 could not be described as one leadership style. The leader of this team applied different leadership styles such as transformational, heroic, as well as post-heroic leadership styles. In this team there was an accountable leader who had superior knowledge; all wisdom was concentrated in him, and he displayed a sense of power and confidence. The leader knew what actions to take during the game, although he has not played that game before. He possessed superior knowledge which helped him to make the right decisions. The dialog below is an example of participant 2 (the leader's) omnipotence:

Participant 1: 'Do we need to train (workers)?'
Participant 2: *'Yes, but it is too expensive.'* (He is upgrading houses)
Participant 1: 'How about (if) we build this one?'
Participant 2: *'Maybe not... we have to make more money.'*

Participant 2 knew what to do in this situation, and he determined his teammates' opinions to be less valuable at this point. In the dialog above, even if participant 1's opinions were turned down by the leader, he agreed with it without insisting on his opinions and without trying to persuade participant 2, which could suggest that participant 2 (the leader) was respected for his knowledge and rightness. This behavior could be described as an idealized influence which is a characteristic of transformational leadership – leaders are admired, respected and trusted. At the same time, the leader of this team applied intellectual stimulation to manage his team. During the competition, participant 2 was the leader who controlled the mouse, read aloud game instructions, and tried to generalize the guidelines for the team. There were several times when two other team members brought up questions, and he answered them very quickly. Although he controlled the mouse during the whole process, he often asked his team members opinions before implementing ideas; hence, team members were included in the process of addressing problems and finding solutions. They participated in making some of the decisions, such as what to do next, but they depended on their leader to make the final decision. Thus, the leadership style was democratic and leader-centered at the same time. Group participation and discussion was encouraged by the leader. Below is an example:

Participant 2: 'Do you want a workshop?, then, he explains, 'the strategy of this level is to build a house and get cash, and you need to build a library.'
Participant 3: 'We need more training.'
Participant 2: 'Oh, mail! That's good.' (He opens the mail, team members look at the screen), then he asks: 'how do you feel about an upgrade?'

From the dialog above, it could be noted that participant 2 asked for his teammates' opinions because probably he wanted to share the responsibility. Thus, the leadership style present in this team could be interpreted as shared leadership also. The leader shared his leadership to some degree with the team members; he asked them what they think should be done next and waited for their agreement to proceed further. However, this kind of leadership does not always have a good influence on team performance. Sometimes leaders cannot make any decision without getting input from their team members. This slows down the process and the organization cannot react quickly enough. In this team, participant 2 was often asking his team members for opinions (although not all of their ideas were adopted and implemented) that helped him to lead the team to win the game. The best leaders favor participation, but also know when they need to be directive or make decisions on their own. According to Lingham (2004), teams with high dependence on a single leader had lower performance, less member satisfaction, and a decreased climate of psychological safety. The dependence on a single leader could be observed in Team 1; however, in the case of this team, team members did not show the aforementioned attributes. Overall, there was no conflict in the team during the playing the game. Everyone seemed to be content with the decisions participant 2 made, and the atmosphere of the team was harmonious. Team members were satisfied with the leadership styles the leader implemented. In summary, leadership techniques used by the leader brought success to the overall team performance; this team won the game competition.

### Team 2.

This team's leadership style could be described as heroic leadership and to some degree as a post-heroic leadership. Disagreements in this team came from the fact that one participant dominated and was pushing other team members to realize his ideas. It could be noted from observations of this team that he was behaving like a heroic-autocratic leader. He issued orders and he expected them to be followed without questions. He wanted to determine all activity tasks and steps. A high degree of control is a main characteristic of the autocratic leadership style. He wanted to win the game and he was concentrated more on the game performance than on his team members' satisfaction and their opinions. The example of participant 4's autocratic leadership style can be noted in the dialog below:

Participant 5: 'What should we do?'
Participant 4: 'We buy houses later, buy one Tudor, we can sell it later.'
Participant 6: 'Colonial first.' (He meant colonial buildings)

Participant 4: ‘No, definitively don’t do it, buy Tudor. Better to build the Tudor. *Listen to me!*’
(Participant 6 was against that)
Participant 4: ‘Sell a Tudor and buy this one of a higher price, *listen to me!’*

In addition, coercive leadership styles could be noted from his behavior as he demanded immediate compliance. This leadership style is the least effective. It is appropriate during an emergency or crisis situation, and when used in the short term, but it is not successful in the long term. In this leadership style, team members do what they are told because they fear being reprimanded. As a result, team members are easy to manage and the leaders have all the control. Unfortunately, team members feel burdened to question ideas and make suggestions, which prevent good ideas from being heard. Another downside involves team members never deviating from their leaders’ orders, even if they do not agree with the orders. When participant 4 was behaving like a demanding leader who gives orders to team members, they disliked it. During the game, they were willing to collaborate with each other more, but not with participant 4. They were not satisfied with participant 4’s attitude towards them and the game competition. However on the contrary, highlights of shared leadership style could be noticed from this team conversations. From time to time there was mutual influence between the team members. In particular, the female participant was asking two other members’ opinions and encouraged them to exchange ideas. She had the mouse in her hand during the game’s competition, and she was trying to discuss with others all decisions she was about to take. The two other participants were expressing their opinions to her, and she was implementing what she had been told to do. In general, all team members tried their best to win the game, and ideas were coming from every team member. They always tried to help each other by expressing opinions and comments. Here is a note from the observations showing highlights of shared leadership style and mutual influence in this team: participants 4 and 5 had an idea of buying a house, and participant 6 commented as follows:

Participant 6: ‘Not buy, let’s build a house.’
(They listened to his comment and constructed the house)
Participant 4: ‘We can sell it later.’
Participant 6: ‘No, let’s sell it now.’
Participant 5: ‘We’ll sell it when we don’t have money...’

Based on the results of the study, the mix of leadership styles implemented in this team cannot lead the team to be successful and to achieve high performance. Team 2 was the second team to finish the game, and perhaps this was due to one participant’s coercive leadership style which was not effectively implemented within the team. In addition, team members were not satisfied with participant 4’s dominant influence.

### Team 3.

In this team, no clear leadership style could be distinguished, however collaboration could be noted from the team’s dialogs but nobody in this team wanted to have a leadership responsibility nor did anyone have a personality with which to influence others. Leadership involves influence; without influence, leadership does not exist. This team’s way of playing the game lacks clearly distinguished leadership style, although highlights of shared leadership could be noted in the team’s conversations. This team’s game score was the lowest, so they lost the game competition. In general, the lack of leadership could be explained by the “official leader” being not present – physically or mentally. Most of the time, team members want to have their leaders to direct, inform, or give them feedback. Thus, when the team has no leader, the team’s potential is hindered. In addition, most team members want to feel a connection to their team; however, without the leader, there is no organization, but rather chaos. See the note from the observations below as an example of Team 3’s lack of organization:

Participant 8: ‘We have to build it. How many do we have to build? 75? We’re out of materials. So what’s next? Do we have additional instructions? Or just make more money?’
Participant 7: ‘We keep making money.’
Participant 8: ‘Our strategy is to make money, right?’ (Silence, nobody answered the question)

From the dialog above, one could notice uncertainty of what they have to do. Later, during the game, when one team member was trying to organize the structure of the team, other team members did not consider that to be necessary. See the dialog below as an example:

Participant 7: ‘We have to buy a house. Raise income.’ (Spoke louder)
(Team members moved forward to read the game instructions)
Participant 8: ‘We don’t have enough money’.
Participant 9: ‘Materials.’
Participant 8: *‘OK, now each of us should be in charge of one job.’*
Participants 7 and 9 laughed and nodded their heads.

(They almost developed the strategy but without a success)
Participant 7: 'Buy a house to raise income. We need to buy or construct.' Then he adds: 'Time is running out…'

In this team, nobody had authority or an answer to the questions other team members were asking. Furthermore, nobody wanted to be responsible and everyone was making suggestions. Nobody demonstrated a tendency to dominate this team and to be a leader. Also, nobody's opinion was respected or considered to be important, they avoided getting involved when important issues aroused. They usually made a decision based on a situation rather than a person. This team did not respond to situations and problems systematically. The team cannot have successful performance without a leader who has the authority to make decisions. A team needs a leader to organize its structure and to successfully lead team members in a complex environment. This team leadership could be described as laissez-faire and team's behavior could be described as passive/avoidant behavior. The "passive leaders" avoid specifying agreements, clarifying expectations, and providing goals and standards to be achieved by followers. This style has a negative effect on desired outcomes and has negative impacts on followers and associates. Therefore, the lack of a leadership role in this team probably caused them to lose the game competition and have the lowest score. These results provide added support for prior research that has shown the use of managing-by-exception (laissez-faire) to be an ineffective leadership style (Avolio, Waldman, & Einstein, 1988).

## Discussion

The main concern of the study was to research how leadership styles will affect the team performance during a computer game training. As a result of the game competition, Team 1 won the game. The leader of this team used different leadership styles such as transformational: shared and democratic leadership styles. The best leaders do not just use one style of leadership – they are skilled at several, and have the flexibility to switch between styles as the circumstances dictate (Goleman, 2000). In the case of Team 1, the transformational leadership and the dependence on a single leader (leader-centered style) resulted in effective team performance. However, the leader-centered leadership style is hardly used alone in today's organizations. One leader alone is not enough in today's complex work environment. It is becoming more difficult for any one person to be an expert in all aspects of the work that needs to be done. In fact, research indicates that poor-performing teams tend to be dominated by the team leader, while high-performing teams display more dispersed leadership patterns, i.e., shared leadership (Pearce, 2004). Shared leadership is a complex process, but it was successfully implemented in Team 1. In this team, the role of the heroic leadership was implemented with shared and democratic leadership styles that made a positive impact on the team performance. On the contrary, in Team 2 shared leadership implemented with autocratic and coercive leadership styles did not bring effectiveness to the team performance. The leader's personality was too strong and his demanding attitude towards other team members had a negative influence on the team members and the team performance. He wanted to direct and control all decisions without any meaningful participation by the team members. With this leadership style, it has been found that most of the decisions made lack in creativity. Although shared leadership was applied in Team 2, it did not help to win the game. In Team 3, there was no clearly distinguished leadership style which probably caused them to lose the game. Team members demonstrated a failure to take responsibility for managing and decision-making. No one in this team provided direction or support; rather, they showed a lack of caring for their team's performance. This 'non-leadership' style could be described as laissez-faire leadership, a principle which emphasizes independence. So, team members are left alone to do their work with little direction or supervision. The conclusion from these results is that leadership is needed for successful team performance.

According to Brousseau, Driver, Hourihan, & Larsson (2006), in order to advance in complex organizations, leaders need to learn and apply more advanced decision styles for using information and evaluating options. It was found that leader decision styles must evolve to include more complex information processing that are capable of managing the challenges of a turbulent environment. Today's organizations are more complex than they used to be, and traditional, heroic leadership styles, when used alone, are not enough to be successful. Therefore, teaching leadership should change as well. One approach to teach leadership is by playing a strategic computer game in teams. Teaching with computer games could prepare students to better cope with business world complexity. Gaming sessions as described in the paper could be implemented in college business education to teach students which leadership styles should be used in a specific situation to manage similar situations in their future workplaces. Classes based on organizing training sessions with strategic computer games to teach leadership styles have two advantages. First, students experience how they behave in pressure situations (competition, time pressure, etc); these situations push them to be more successful and creative or they cause them to drop out without caring about actively participating in the training session. Second, they see which team wins the game, attempt to interpret the game results, and come to the conclusions as to why their team performance was not effective if their team lost. This type of experience teaches students how to behave in certain situations and how the competition affects their behavior. After the gaming session, they will realize from the experience which type of leadership styles work well and how to behave next time when placed in a similar situation. However, not everyone has a predisposition to being a leader; a particular personality type is usually needed. People who have a 'leader personality' could be trained to be better leaders and training such as the type mentioned in this study could help them to develop strategies on how to be more successful. In summary, playing a strategic computer game in teams of three could provide the tool to learn applying different leadership styles in different situations.

The interesting area for future inquiry would be to organize leadership training sessions in which participants play in more than one team. In this way, they will have an opportunity to experience different leadership styles with different team members. It would be fruitful for the game participants to conduct interviews with them after the training and evaluate the leadership styles they have experienced. This kind of debriefing could help participants to articulate what they have learned for the purpose of knowledge building (Jonassen, Peck, & Wilson, 1999).

## References


Alvesson, M., & Sveningsson, S. (2003). 'Managers doing leadership: The extra-ordinarization of the mundane'. *Human Relations, 56*(12), 1435-1459.

Avolio, B. J., Waldman D. A., & Eisten W. O. (1988). Transformational leadership in a management game simulation. Impacting the bottom line. *Group & Organization Studies, 13*(1), 59-80.

Baker, E. L. & O'Neil, H. F., Jr. (2002). Measuring problem solving in computer environments: current and future states. *Computers in Human Behavior, 18,* 609-622.

Bass, B. M. (1985). *Leadership and performance beyond expectations.* New York: Free Press.

Bass, B. M. (2000). The future of leadership in learning organizations. *The Journal of Leadership and Organizational Studies, 7*(3), 18-40.

Bass, B. M., & Avolio, B. (1994). *Improving organizational effectiveness through transformational leadership.* Thousand Oaks, CA: Sage.

Bradford, D. L., & Cohen, A. R. (1998). *Power up: Transforming organizations through shared leadership.* Chichester, England: Wiley.

Brousseau, K., Driver, M., Hourihan, G., & Larsson, R. (2006). The seasoned executive's decision-making style. *Harvard Business Review, 84*(2), 111-121.

Burke, C. S., Stagl, K. C., Klein, C., Goodwin, F. G., Salas, E., & Halpin, M. S. (2006). What type of leadership behaviors are functional in teams? A meta-analysis. *Leadership Quarterly*, *17,* 288-307.

Bryman, A. (1996). Leadership in organizations. In S. R. Clegg, C. Hardy & W. R. Nord (Eds.), *Handbook of organization studies* (pp. 276-292). London: Sage.

Conger, J. A., Kanungo, B. N., & Menon, S. T. (2000). Charismatic leadership and follower effects. *Journal of Organizational Behavior, 21,* 747-767.

Day, D. V. (2001). Leadership development: A review in context. *The Leadership Quarterly*, *11,* 581-613.

Edmondson, A. C., Bohmer, R. M., & Pisano, G. P. (2001). Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative Science Quarterly, 46,* 685-716.

Eicher, J. P. (2006). *Post-heroic leadership: Managing the virtual organization.* Retrieved October 20, 2009, from http://www.pignc-ispi.com/articles/management/post-heroic.htm

Ensley, M. D., Pearson, A., & Pearce, C. L. (2003). Top management team process, shared leadership, and new venture performance: A theoretical model and research agenda. *Human Research Management Review, 13,* 329-346.

Fletcher, J. K. (2004). The paradox of postheroic leadership: An essay on gender, power, and transformational change. *Leadership Quarterly, 15,* 647-661.

Gemmill, G., & Oakley, J. (1992). Leadership: An alienating social myth? *Human Relations, 45*(2), 113-129.

Goleman, D. (2000). Leadership that gets results. *Harvard Business Review*, *78*(2), 78-90.

Hartenian, L. S., Schellenger, M.,&Frederickson, P. (2001). Creation and assessment of an integrated business course: One college's experience. *Journal of Education for Business*, *76*(3), 149-159.

Hense, J., & Mandl, H. (2009, August). *Learning with digital games: Potentials and best practice in the light of theories of learning, emotion, and motivation.* Paper presented at the 13th EARLI Conference, Amsterdam.

Howell, J. P., & Costley, D. L. (2001). *Understanding behaviors for effective leadership.* Upper Saddle River, NJ: Prentice Hall.

Jenkins, H. (2002). Game Theory: Digital Renaissance, *Technology Review. MIT.* March 29, 2002.

Jonassen, D. H., Peck, K. L., & Wilson, B. G. (1999). *Learning with Technology: A Constructivist Perspective.* Upper Saddle River, NJ: Merrill/Prentice Hall.

Kafai, Y. B. (2006). Playing and making games for learning: Instructionist and constructionist perspectives for game studies. *Games and Culture, 1,* 36-40.

Kark, R., Shamir, B., & Chen, G. (2003). The two faces of transformational leadership: Empowerment and dependency. *Journal of Applied Psychology*, *88,* 246-255.

Katzenbach, J. R. (1997). The myth of top management team. *Harvard Business Review, 75*(6), 82-91.

Kiili, K., & Lainema, T. (2008). Foundation for Measuring Engagement in Educational Games. *Journal of Interactive Learning Research*, *19*, 469-488.

Lehtinen, E. (2002). Commentary. Developing models for distributed problem-based learning: Theoretical and methodological reflection. *Distance Education*, *23*(1), 109-117.

Light, R. J. (2001). *Making the most of college: Students speak their minds*. Cambridge, MA: Harvard University Press.

Lingham, T. (2004). *Developing a measure of conversational learning spaces in teams*. Unpublished doctoral dissertation, Case Western Reserve University, Cleveland, OH.

Mello, J. A. (2003). Profiles in leadership: Enhancing learning through model and theory building. *Journal of Management Education, 27*(3), 344-361.

Michaelsen, L., Bauman Knight, A., & Fink, L. D. (2004). *Team-based learning: A transformative use of small groups in college teaching.* Sterling, VA: Stylus.

Oblinger, D. (2004). The next generation of educational engagement. *Journal of Interactive Media in Education, 8*, 1-18.

O'Neil, H. F., Jr., & Andrews, D. H. (2000). *Aircrew training and assessment.* Mahwah, NJ: Lawerence Erlbaum Associates, Inc.

O'Neil, H. F., Jr., & Fisher, Y. C. (2004). A technology to support Leader Development: Computer Games. In D. V. Day, S. J. Zaccaro, & S. M. Halpin (Eds.), *Leader development for transforming organizations: growing leaders for tomorrow* (pp. 99-121). Mahwah, NJ: Lawerence Erlbaum Associates, Inc.

Pearce, C. L. (2004). The future of leadership: Combining vertical and shared leadership to transform knowledge work. *Academy of Management Executive, 18*(1), 47-57.

Perry, E. L., & Ballou, D. J. (1997). The role of work, play, and fun in microcomputer software training. *ACM SIGMIS Database, 28*(2), 93-112.

Pescosolido, A. T. (2001). Informal leaders and the development of group efficacy. *Small Group Research, 32,* 74-93.

Rost, J. (1991). *Leadership for the twenty-first century.* Westport, CT: Praeger.

Silverman, D. (2006). *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction* (3rd ed.). London: Sage.

Stephen, J., Parente, D. H., & Brown, R. C. (2002). Seeing the forest and the trees: Balancing functional and integrative knowledge using large-scale simulations in capstone business strategy classes. *Journal of Management Education*, *26*(2), 164-193.

Yukl, G. (1989). Managerial leadership: A review of theory and research. *Journal of Management, 15,* 251-289.

Yukl, G. (1998). *Leadership in organizations* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Yun S., Cox J., Sims, H. P., Jr., & Salam S. (2007). Leadership and Teamwork: The Effects of Leadership and Job Satisfaction on Team Citizenship. *International Journal of Leadership Studies, 2*(3), 171-193.

Zaccaro, S. J., Rittman, A. L. & Marks, A. M (2001). Team leadership. *The Leadership Quarterly, 12,* 451-483.

## Appendix

Build-a-lot – a strategic computer game, where players are real estate moguls whose task is to take over the housing market as they build, upgrade, sell and buy houses for huge profits. The objective of the game is to get Net Value to the highest possible number by building, upgrading and selling properties. The team with the most profit and the highest achieved level wins the game. If the players manage to get every operation correct, they pass the current level in the game and advance to the next level. If a team fails a level in the game, they have to repeat that level from the beginning. During the game, players get the instructions from the mayor of the city in which they are building houses. To construct a property, players must have the blueprint, enough workers and enough materials. The players' task is to move from one town to another to improve them and achieve all the goals in a given period of time. They are asked to build different kinds of houses and buildings, earn a specific rental income, or earn cash. To achieve these goals, players have to choose exactly what they have to build and demolish, which requires certain calculations. Players can build the following kinds of houses: Rambler, Colonial, Tudor, Estate, Mansion, and Castle. They should buy the blueprint of each of these, the materials needed, and a certain number of workers. They can also build a number of buildings like a Post office, Library, Workshop, Bank, Sawmill, etc. The main advantage of this game is the fact that it makes players think and calculate items (for example, taxes that they should pay) in order to win. In this environment, participants played in teams, with three members on each team. Game play with small teams forces team members to express and exchange their ideas and opinions to form a strategy. Furthermore, teams are indispensable for successful performance, and individuals have to be trained to be able to work efficiently and effectively in groups.

# Validation of a Learning Progression: Relating Empirical Data to Theory

**Abstract:** Learning progressions (LPs) are theoretical models of learning trajectories in a domain. Recent policy reports have touted LPs as a promising approach to align standards, curriculum, and assessment. However, the scholarship on LPs is sparse, and the jury is still out on the theoretical and practical value of this approach. To realize any potential of LPs we need to systematically validate and refine these hypothetical models in real-world contexts. Such validation efforts are challenging, as they require the coordination of messy empirical data with, often, under-specified theoretical models. In this paper we report on our efforts to validate a genetics LP through a two-year longitudinal study in middle school. We describe how we used interview data to refine the hypothesized levels of progression in our LP and to identify contingencies between constructs (big ideas) within our LP. We conclude with some tentative heuristics for coordinating data and LP models.

## Introduction

Recently there has been a resurgence of interest in learning progressions (LPs) - a theoretically driven pedagogical approach that emphasizes the learning of big ideas and practices, in a domain, over extended periods of time (Author et al., 2009a; Lehrer & Schauble, 2009). Advocated by policy reports as means to align standards, curriculum, and assessment, LPs have the potential to be a transformative new approach to science education reform (e.g. NRC, 2007). However, the scholarship about LPs is still in its infancy and extensive validation efforts are needed to move this scholarship forward and ascertain the theoretical soundness of LPs and practicality of using them in the science classroom.

LPs are theoretical models of learning in a domain that describe progressively more sophisticated levels of reasoning that a learner exhibits as she develops expertise (NRC, 2007). These progressions describe learning as it unfolds over extended periods of time (grades and grade bands). LPs are not developmentally inevitable but require targeted instruction and curriculum. Most importantly, LPs are hypothetical constructs; and while they are grounded in research in student learning, to the extent possible, they do require validation. Existing examples of progressions vary in the number, and hence grain size, of the levels they postulate (e.g. Author et al., 2008; Schwarz et al., 2009; Songer, Kelsey & Gotwals, 2009). These differences raise the question of what is a useful grain size for LPs? Too many levels and it is difficult to generalize (you loose the forest for all the trees), too few levels and one looses explanatory power (Lehrer & Schauble, 2009). Determining the appropriate number of levels is not merely a theoretical exercise; it also entails empirical tests (often classroom studies) in which LPs can be validated. Such validation is not trivial as students' ideas, or levels of performance, may not fall neatly into the LP's expected levels. The researcher must decide, given the empirical data, when to add new levels, consolidate levels, or remove levels from the hypothetical LP.

LPs also differ in the number of big ideas or constructs they encompass, and in the ways in which progress along one construct is related (or not) to progress along other constructs. Wilson (2009) described different assessment structures of LPs that exemplified distinct types of relationships between constructs. For example, progress along two constructs may occur independently in a parallel structure, or constructs may be linked such that progress along one construct may only begin after the attainment of some level of sophistication on another. Establishing the existence of such links is an important part of the LP validation process. Like the validation of the number of levels in an LP, the validation of links between constructs presents both theoretical and methodological challenges. In this paper we describe out efforts to address the challenge of validating an LP, and coordinating between the hypothetical model and empirical data, in the context of a genetics LP previously described (Author et al., 2009a). We describe how we refined the number of levels (for a construct in our LP) and how we identified contingencies, or links, between levels of two constructs within the LP.

Here we will focus on two constructs within this eight construct LP and describe the progression along these constructs as hypothesized in the LP. We chose to focus on these constructs as they are core ideas molecular genetics that are often underrepresented in middle school curricula, and therefore students had a wide array of understandings about these ideas (providing a fruitful context for our validation efforts). The first construct (construct B of our LP) concerns the nature of the genetic information. Genes are essentially instructions that encode the order and type of building blocks of proteins. While the outcomes of genes have far reaching effects, the information itself mostly specifies the structure of one, very important, biological molecule - the protein (Author et al., 2009a). The second construct (construct C of our LP) focuses on the idea that proteins act as the intermediary between genes and traits and are required for the functioning of all organisms (Author et al., 2009a). These molecules have many biological functions such as catalyzing chemical reactions, building channels, structural support, relaying messages within cells, among other functions. Understanding the link between genes and proteins, and the role of proteins in biological traits is at the core of molecular genetics.

Table 1 illustrates the levels of knowledge to be reached by students at the specified grades. We explain the hypothetical progression for these constructs in more detail in the results and discussion section.

Table 1: Construct B and C from the hypothetical learning progression.

| | Level 1: Grades 5-6 | Level 2: Grades 7-8 | Level 3: Grades 9-10 |
|---|---|---|---|
| Hypothetical Construct B (Author et al., 2009a) | Genes provide instructions that determine how organisms develop. | Genetic instructions encode for proteins which have specific functions within organisms. | Genes encode for amino acids, which make up proteins. |
| Hypothetical Construct C (Author et al., 2009a) | Cells are one level of organization within our bodies. Cells have specific organelles that help the cell perform its function. | Proteins perform specific functions within cells. Genetic mutations can result in changes within the structure and function of proteins. | The amino acid sequence of a protein determines its shape/function. There are different kinds of genetic mutations which can affect the structure and function of proteins. |

## Methods

### Study Context

The data analyzed in this paper was collected as part of a two-year longitudinal study designed to empirically test the genetics learning progression described by Author et al. (2008). Due to logistical concerns, 8th graders were used as a proxy for high school students. In order to validate the progression we instantiated it within two instructional units. These units were project-based, where students investigated three genetic diseases in the first unit and focused on a genetic engineering problem in the second unit. These units were developed by a collaborative design team of educational researchers and teachers. The units were implemented by of one of the teachers on the design team in her mixed grade classroom (6th, 7th, and 8th grade students) in a small urban, public K-8 charter school. The school has a total of 234 students (with 59 6-8th grade students in the study) with a large proportion of minority students and about 35% of the students eligible for free or reduced lunch.

### Data Collected

Four types of data were collected: (a) pre and post written assessments, (b) artifacts of student work (including a performance assessment), (c) individual pre and post clinical interviews, and (d) video tapes of classroom instruction and group work. For this paper we discuss the collection and analysis of the clinical interviews only. We conducted pre and post interviews during the first year of the study (before and after implementation of the first unit); and a second round of post interviews were conducted at the end of the implementation of the second unit. The interviewed students (N=24) were selected by the teacher to offer the most representative sample of students from the study classrooms (Total N=60).

Interviews included three open-ended tasks that all provided opportunities for students to discuss their understanding of how genes bring about their effects and the role of proteins in genetic phenomena. In the first task students were asked to reason about the inheritance of a genetic disorder based on a family pedigree and to speculate about the molecular mechanism underlying this disorder. Part of this task specifically focused on how students used ideas about genes and proteins to formulate explanations of the mechanism of the genetic disorder. The second task, involving a genetic trait in bacteria (chemotaxis), prompted students to provide a hypothetical explanation, at the cellular level, how bacteria can sense substances in their environment. As well as to speculate about cellular mechanisms that could render mutated bacteria unable to sense (Author et al., 2007). In the final task we asked students for specific definitions of genetic terminology such as gene, protein, DNA, and chromosome. In this task we also prompted students to share their ideas about how these entities are related to one another.

### Data Analysis

In this paper, we report our analysis of the second round of post interviews. We chose this set of interviews because it exemplified the largest variation in student ideas and represented multiple levels of understandings in the domain. This variation in student ideas provides a suitable context for examining the different levels of sophistication and any relationships between ideas in distinct constructs in the progression. All interview transcripts were blinded for analysis. As noted earlier, we will discuss students' understandings of constructs B and C in the progression, which pertain to the molecular model.

We conducted a content analysis of all interview tasks and characterized students' ideas regarding the two constructs in the progression. The expected levels of sophistication, as detailed in the progression, informed the initial coding schemes. Through an iterative process these initial coding schemes were refined to capture relevant variation in the data. Within the first construct, several levels of sophistication were identified and they ranged from identifying genes as passive particles associated with traits, to the most sophisticated level of understanding- the cellular mechanism by which genes code for proteins. Similarly, several levels of sophistication were observed for the second construct, ranging from a vitalistic view of proteins (a general notion of proteins as important to vitality and health) to a view of proteins as having specific functions that mediate genetic effects in an organism. We also determined the number of student responses associated with each level of sophistication for both constructs.

Another aspect of analysis was to determine dependencies, or links, between the two constructs; we characterized the ways in which students connected ideas relating to the two constructs. For this analysis we examined the interview as a whole and looked for evidence of connections between ideas in the two constructs. We created individual profiles for each student that described the extent to which the constructs were linked. We were able to triangulate our findings by referring back to two previously coded aspects of the first interview task: how gene mutations affect protein function and how students organized ideas about genes and proteins. Referring to this analysis provided supporting evidence for the contingencies determined between the two constructs. Inter-rater reliability for these two aspects was 91%.

## Results and Discussion

### Refining the Levels of a Learning Progression

One of the major difficulties with validating LPs is mapping the theoretical framework onto empirical findings. In our analyses we found several levels of understanding that had not been captured by the hypothetical progression. This was particularly prominent in our analysis of Construct B, which deals with the nature of the genetic information. At the lower anchor of the progression for construct B are non-information based views of the genetic material, namely, genes as passive particles associated with traits. As a result of instruction students are expected to develop information-based views of genes at various levels of sophistication. According to the progression students move from a view of genes as instructions "for how organisms grow, develop, and function", but without a notion of genes as coding "exclusively" for proteins (level 1 of the progression), to a view of genes as productive instructions for protein molecules that perform tasks within cells (level 2), and finally to a view of genes as instructions that encode for the order and type of amino acids in a protein (level 3) (Author et al., 2009a, p.660). Our data did reflect the existence of these levels. However, we also found evidence for other students' ideas and levels of understandings that were not previously captured in the LP.

Thus the data allowed us to refine the construct map regarding idea B in the following ways. First, while the instructional units mapped onto levels 2 and 3 of the progression, we found that students in our study entered the progression with understandings that were below a level 1. The progression at a level 2 anticipates that students will already have an information-based view of genes (essentially level 1 of the LP is getting to this understanding). What we found is that 23 of the 24 students were at a level 0 (non-information based view of genes) upon entering the study and only 1 student still held this view at the end. Given that students did not receive any genetics instruction prior to the study, this finding is not surprising. However, it does underscore the importance of moving students from a non-information based view of genes to an information-based view of genes as a precursor to any of the more sophisticated understandings regarding gene function.

Second, we found that many students struggled with the transition to an information-based view of genes, and that there seems to be an intermediate level of understanding, not captured by the current progression, that instantiates this transitional understanding (genes as active information). Students who held this view believed that genes could literally "tell" proteins, the cell, or the body how to do a function. This view is information based, but no translation is required between the genetic information and a physical entity, as the gene itself directs the function. For example, in the case of a genetically inherited blood clotting disorder, Barry claimed that "[the genes are] giving instructions for uh, to tell the cell, to tell the cell what to do". When we identified this transitory level of understanding we had to decide whether it was sufficiently distinct from other levels of understanding to merit inclusion, as its own level, in the progression. Given that the transition from the non-information based to an information-based view of genes is a particularly challenging one (Author et al., 2007; Venville & Treagust, 1998) we decided that capturing this understanding as a distinct level was merited for two reasons. First, cognitively this understanding reflects an ontological shift in students' understandings of genes, in essence the beginning of an important conceptual change. Second, and following from the first, it is pertinent that teachers (and curriculum/assessment designers) recognize this understanding as bridging across ontologies and leverage it in their instruction.

We also found that there were variations in the ideas that fell within level 1 of the progression (genes as information for growth and development). At this level of the progression, students tend to assume that genes can code for multiple biological entities (and functions) such as cells, tissues, organ functions, and whole traits. While students may include proteins as one of the many things genes code for, they usually do not view proteins as the predominant product of the genetic instructions. We found that some of the students in our study included only proteins and cells as entities specified by the genetic information. We initially thought of separating this understanding out into its own level, as thinking at the cellular level seemed more sophisticated than thinking at the tissue or organism level. However, we decided against this because our sense was that students tended to map the genetic information onto the lowest level of organization they were familiar with, for some students this was organs, for others cells. Making a distinction between these ideas did not seem like a true representation of students' knowledge of genes, but rather their knowledge of biological organization levels. Moreover, we felt that this understanding tended to be context dependent such that a student may map the genetic information onto cells in one context and onto organs in another.

Table 2 illustrates our data-driven refinements of construct B. Overall, we found that 6 of the students were able to reach level 2 of construct B of the progression. There were 5 students still at level 1 and 9 students were at the new transitory level between levels 1 and 2 of the progression. Very few students were able to reach level 3 of the LP (4 students total), and even for those who did express a level 3 understanding it was not as robust as we had hoped.

Table 2: Data-driven refinements of construct B.

| | Level 1: Grades 5-6 | Level 2: Grades 7-8 | Level 3: Grades 9-10 |
|---|---|---|---|
| Hypothetical Construct B (Author et al., 2009a, p. 660) | Genes provide instructions that determine how organisms develop. | Genetic instructions encode for proteins which have specific functions within organisms. | Genes encode for amino acids, which make up proteins. |

| New LP level | | Original LP level | Level Description |
|---|---|---|---|
| Construct B Revised | Level 1 | Level 0 (lower anchor) | Genes are non-informational in nature. They are passive particles associated with traits. |
| | Level 2 | Was not captured | Genes are non-informational in nature. They are active particles associated with traits. Genes "determine" traits. |
| | Level 3 | Was not captured | Genes are active instructions that "tell" proteins, the cell, or the body to carry out specific functions. |
| | Level 4 | Level 1 | Genes have information about biological entities and function at multiple organization levels. |
| | Level 5 | Level 2 | Genes are instructions for molecules (many of which are proteins) that carry out functions within the organism. All organisms use the same genetic language for their instructions. |
| | Level 6 | Level 3 | The genetic code is translated into a sequence of amino acids that makes up the protein. Almost all organisms use the same genetic code. |

We now turn to our findings regarding construct C of the LP, which involves the understanding that proteins have a central role in organism function and are essentially the link between the genetic information (genotype) and the resulting physical trait (phenotype) (Author et al., 2009a). Again, three levels were defined in the hypothetical LP. The first level involves a non-protein based understanding of biological phenomena whereby proteins are not yet central to biological function (and may not be mentioned by students) however, students are able to reason about biological organization levels and in particular cells as the building blocks of life and as having specific functions within tissues and organs (see Table 3). Level 2 of the progression anticipates a protein-based view of biological function; at this level students should be able to explain biological phenomena by invoking proteins as central players, and be able to propose specific functions for those proteins in cells and tissues. Finally, by the end of the LP (level 3) students are expected to understand that that proteins are made up of amino acids and that the properties of the amino acids determines the proteins' three dimensional structure, which, in turn, determines the protein's function. At this level students should also be able to reason about the ways in which changes to the genetic code (mutations) can affect the structure and thus the function of a protein and ultimately the trait.

For this construct, we analyzed the interviews and identified three distinct levels of sophistication that students exhibited with respect to proteins (see Table 3). In the first level students had vitalistic notions of proteins and considered proteins as solely providing positive health benefits for the body. For example, Alyssa suggested that "proteins, they do like, um, help you grow, with like your bones and stuff", reflecting a vitalistic view of proteins. These students were unable to suggest specific protein functions but did have a sense that proteins are necessary. These students also explained that without proteins negative health affects would ensue, that is, they realized that proteins have an important biological role and that there are consequences to interfering with protein function. This level was not captured by our hypothetical LP, but is considered to be below level 2 of the hypothetical LP in terms of sophistication. In the next level students explicitly described proteins as central to genetic phenomena but were unable to provide any contextual examples, which corresponds to level 2 of the hypothetical progression. These students held a schema which invoked "proteins" as an explanatory element of the genetic phenomena. That is, these students knew that if a phenomenon is genetic, it must involve proteins. However, they struggled to apply this schema in context and provide specific examples of protein involvement. For example, when presented with a reasoning task concerning an inherited genetic disorder, Daniel immediately discussed inheritance of genes from family members. He also agreed that proteins are involved in the genetic disorder, but when prompted to give an explanation of this he responds that "like [genes] gives protein to the person, so the person can be healthy and stuff like that". Daniel believes that proteins are involved in the genetic phenomena but cannot provide examples beyond his acknowledgment of their presence. In the final level students could instantiate their ideas about proteins as related to genetic phenomena by providing specific examples of protein functions and how a genetic mutation might affect protein function (corresponding to level 3 of the LP).

Table 3 illustrates our data-driven refinements for construct C. We did not have any students at a level 1 of the hypothetical progression by the end of the two units. However, we did have evidence from earlier interviews that a non-protein based view of genetic phenomena exists. We found that 12 students remained at level 1.5 of the refined construct C. Several students (8 total) were able to reach a level 2 understanding where some could provide contextual examples, while others could not (suggesting this level could be split further). Very few students (4 total) were able to reach level 3 of the refined construct C and could reason about amino acids as building blocks for proteins. Again, as with construct B, their ideas about proteins were not as robust as we had hoped.

Table 3: Data-driven refinements of construct C.

| | Level 1: Grades 5-6 | Level 2: Grades 7-8 | Level 3: Grades 9-10 |
|---|---|---|---|
| Hypothetical Construct C (Author et al., 2009a) | Cells are one level of organization within our bodies. Cells have specific organelles that help the cell perform its function. | Proteins perform specific functions within cells. Genetic mutations can result in changes within the structure and function of proteins. | The amino acid sequence of a protein determines its shape/function. There are different kinds of genetic mutations which can affect the structure and function of proteins. |

| | New LP level | Original LP level | Level Description |
|---|---|---|---|
| Construct C Revised | Was not captured | Level 1 | Cells are one of the basic levels of organization in the body. They can perform specific functions. |
| | Level 1.5 | Was not captured | Proteins are vitalistic in nature. They are good for you and provide positive health benefits without them, general health declines. |
| | Level 2 | Level 2 | Proteins are required for optimum health. Their function is dependent upon their structure. Students cannot provide contextual examples of protein function. |
| | Level 3 | Level 3 | Proteins are central in the genetic phenomena. Their function is dependent upon their structure. Students provide specific contextual examples of how protein function (or lack of function) contributes to the genetic phenomena. |

## Identifying Connections and Contingencies between the Constructs

The theoretical LP in genetics did not include any hypotheses regarding links or dependencies between constructs, there was simply not enough evidence in the literature to support such assertions. Our second goal for this analysis was to ascertain the ways in which understandings along constructs B and C might be related, or dependent upon each other. From a canonical point of view these constructs must be related as construct B essentially defines what biological entity the genetic information specifies (genes code for proteins), and construct C defines the role of proteins in genetic phenomena (the idea that proteins bring about the phenotype). However, it is not clear whether these understandings develop in parallel, or whether a student needs to attain a certain level of sophistication for construct B in order to reason at particular levels for construct C (or vice versa). Towards this end we tried to map out the connections between these constructs as manifested in students thinking. Overall, we found that students had difficulty drawing a relationship between genes and proteins and thus these constructs seemed to exist (and to some extent develop) in parallel.

We found that for 8 of the students there did not seem to be any connection at all between genes and proteins (see Figure 1 – light shading). There were 11 students that demonstrated understandings that suggested a weak link between these constructs, a sort of touch-and-go connection, in which the extent of linkage varied with the task context (see Figure 1 – medium shading). Only 5 students expressed understandings that suggested fairly stable and robust connections between the constructs. Figure 1 further illustrates the distribution of students in each category of connections between the two constructs as well as students' level of understanding for each of the constructs individually. We found that only students who showed a level 6 understanding for construct B (understands that genes provide instructions to make proteins) and a level 3 understanding for construct C (provides contextual examples of protein's role in genetic phenomena) exhibited stable and robust connections between these constructs. For these students, the constructs seemed intertwined and they were able to easily reason about the relationship between genes and proteins. For example, in reference to a genetically inherited skin elasticity disorder, William immediately suggested that "the gene coded for the wrong amino acid so the protein was like misshaped, and then that means that [the protein] didn't work".

| Role of proteins in genetic phenomena (levels 1 – 3) | Genes and their relationship to proteins (levels 1 – 6) | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | | | | | | |
| 1.5 | 3 | 2 | 3 | 2 | 2 | |
| 2 | | | 2 | 2 | 3 | |
| 3 | | | | | | 5 |

Figure 1. Contingencies between constructs B and C.

Not all cases for intertwined understandings were as robust as William's. In another case for intertwined understandings, Amy also held a level 6 understanding for construct B (genes encode for amino acids which provide structure and function for proteins), and a level 3 understanding for construct C (proteins as central to genetic phenomena with contextual explanations). However her understanding of the connection between genes and proteins was not as straight forward. It is also possible that she held a higher level of understanding but did not provide evidence of this in her interview. For example, in the following exchange, Amy was asked to reason about a genetically inherited blood clotting disorder:

Interviewer: Now do you think proteins have anything to do with this?
Amy: Mmm hmm (agrees)
Interviewer: Okay, what do you think proteins are doing?
Amy: Um, I think the proteins that should be doing its job is, has to um, has to send a message to the cells to like, um, to um, to duplicate their cells and um close the, close the, the cut.
Interviewer: Okay, very good. What about genes? Do genes have anything to do with this?
Amy: Mmm hmm (agrees)
Interviewer: Okay how do genes work?
Amy: The genes have instructions to make a certain protein and if there's something wrong with it or they have a mistake, the protein won't work that how it's supposed to work. Or sometimes it doesn't, it isn't there at all.
Interviewer: Okay, very good. So what's wrong with the genes specifically?
Amy: The genes isn't giving the information to…there's something wrong with the instructions and the protein won't give the message to the cells to repair the, the cut or something.

In this example, Amy provides intertwined understandings of the two constructs. She is able to define the role of the protein (causing the cell to undergo mitosis and close the cut) and the role of the gene (to provide instructions to make the protein). She also demonstrates her knowledge that a mutation ("something wrong") in the gene leads to the protein's loss of function resulting in the genetic phenomena. However since she uses ambiguous phrases like "something wrong" and does not explicitly define the role of the protein other than sending messages, this suggests that intertwining the progressions is a progressive process.

An understanding of the role of protein structure and function in genetic phenomena was critical for students to fully explain how genes and proteins are related to one another. This was also reflected in the other two categories of relationships between the constructs (no link, and weak link). If students were at a level 2 on construct C (proteins as central to genetic phenomena without specific contextual examples), they were limited to providing only specific touch-and-go connections between genes and proteins. For example, Kaitlyn (at a level 2 for construct C and level 3 for construct B) was asked how genes and proteins are involved in an inherited skin elasticity disorder:

| | |
|---|---|
| Kaitlyn: | I think it has to do with that their body is not making that protein. |
| Interviewer: | Okay, why isn't their body making that protein? |
| Kaitlyn: | Maybe because there is a mutation in the code or something. |
| Interviewer: | So why would a mutation in the code prevent them from making a protein? |
| Kaitlyn: | Because it's messing everything up and not letting it do its job. |
| Interviewer: | So the mutation is messing up the protein? |
| Kaitlyn: | Mmm hmm (agrees) |
| Interviewer: | So the protein can't do its job. What do we think the job of the protein might be here? |
| Kaitlyn: | Um, to make [the patient's] skin firm and flexible and not like get bruised easily. |

In this example, Kaitlyn was unable to explain the link between a change to the gene and a resultant change to the protein-- what precisely was being "messed up" and how this ultimately led to a missing protein. This example illustrates the incomplete explanations students generate when they only have weakly connected knowledge of the two constructs.

The strength of the linkage was reflected in the level to which they progressed for each construct. For example students reaching level 3 on construct B and a level 2 on construct C (the lowest score possible for linkage) held weaker associations between genes and proteins than a student reaching a level 5 on construct B and a level 2 on construct C. There were several students (8 total) who did not demonstrate any understandings that genes and proteins are related, and this was symptomatic of their understanding of proteins. These students tended to view proteins as vitalistic or being required to maintain optimal health; these views of proteins were paired with views of genes as non-informational or, at most, as active information (level 3 of refined construct B). Two of these students held a level 5 understanding of genes in construct B (genes as information for proteins), but only a level 1.5 understanding of proteins in construct C (proteins as required for optimum health). In these cases, the students both viewed genes as providing information to make proteins, yet at the same time did not view proteins as central to the genetic phenomena (described proteins as required for health and vitalistic). This is unique, since other students who held a similar level 5 understanding for construct B, had at least a level 2 (or higher) understanding for construct C (proteins as central without contextual examples). This suggests that high levels of sophistication for construct B do not necessarily bootstrap connections between the two constructs.

By analyzing the relationships between two constructs within our LP we were able to determine that students' views of proteins were, to some extent, contingent upon their understandings of genes. In many cases students were able to describe genes in terms of genetic inheritance quite accurately (relate genotype to phenotype), but it was not until they viewed proteins as central to genetic phenomena that linkages between ideas about genes and proteins could be made. Based on our data we can both revise these two constructs within our LP and describe the ways in which development along one construct depends, or is dependent upon, development along another.

## Conclusions and Implications

We, like others, are cautiously optimistic about the potential of LPs to inform instruction, curriculum and assessment (Author et al., 2009b; Lehrer & Schauble, 2009; NRC, 2005; NRC 2007). Their potential can only be realized if we can effectively validate and revise them based on empirical findings. Such validation efforts are challenging given the messy, and context dependent, nature of the data. In this paper we presented our efforts to use our empirical findings to inform revisions for a theoretical LP in genetics. Based on this research we next provide a few heuristics and suggestions for the process of refining LPs in relation to data.

Levels should be added (or split) when the ideas inherent to the new level are directly related to the construct and offer explanatory power (in terms of student cognition) or instructional leverage. In the case of our refinement of construct B (Table 2) we added the transition level because it highlighted the initial step of conceptual change. This new level provides explanatory power in terms of modeling conceptual change and is an instructional leverage in the sense that it can be a bridging understanding from a non-information based view of genes to one that is information-based. We did not, however, include a new level to capture students' ideas that genes have information about cells, despite their existence in the data set. This new idea was not directly relevant to the construct as it reflected an understanding of biological organization levels rather than notions about genetic information per se.

While we did not have to consolidate or remove levels based on our data, our sense is that such a move would be necessary if there are no cases, in the data set, of the specific level of understanding, and the level provided little explanatory power in terms of predicting student performance or modeling student cognition. Given that starting point for the LP included only three levels, it was unlikely that we would need to remove any levels. It was the case that we observed very few data points for level 3 understandings, however, we do see this is reason to remove the upper target of the progression. It does, however, suggest ways in which we would need to revise our instructional materials in order to move more students along the progression.

It is very likely, given the complex nature of learning, that movement along one or more constructs in a progression would depend, in some way, on the students' level of understanding of other constructs (Wilson, 2009). In order to identifying such links or dependencies between levels across multiple constructs it is necessary to map out (as we did in Figure 1) the relationship between the constructs as a function of students' level of understanding in each. Such a representation allowed us to see the constraints and affordances that different levels of understandings have in relation to other constructs. In this paper we conducted the analysis for two constructs. Adding more constructs to the mix will certainly present a challenge, but we believe that a similar approach could be used with larger numbers of constructs at play.

We anticipate that our revisions to the genetics LP will be a first step in a series of refinement and revision cycles. In future studies we also intend to vary the curriculum in order to begin to tease apart aspects of student thinking and learning that seem curriculum dependent. This is another challenge in validating the LP approach and much additional work is needed in order to develop our understanding of the relationship between LPs and the instructional context.

## References


Author (2007). The role of domain-specific knowledge in generative reasoning about complicated multileveled phenomena. *Cognition and Instruction, 25*(4), 271 – 336.

Author et al. (2007). Reasoning across ontologically distinct levels: Students' understandings of molecular genetics. *Journal of Research in Science Teaching, 44*(7), 938-959.

Author et al. (2008). Promoting middle school students' understanding of molecular genetics. Paper presented at the *Annual Meeting of the National Association for Research in Science Teaching,* Baltimore, MD.

Author et al. (2009a). A learning progression for deepening students' understandings of modern genetics across the 5th – 10th grades. *Journal of Research in Science Teaching, 46*(6), 655 – 674.

Author et al. (2009b). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching, 46*(6), 606 – 609.

Lehrer, R. & Schauble, L. (2009). Images of learning, images of progress. *Journal of Research in Science Teaching, 46*(6), 731 - 735.

National Assessment of Educational Progress (NAEP). (2006). Science assessment and item specifications for the 2009 National Assessment of Educational Progress. Washington, DC: National Assessment Governing Board.

National Research Council (NRC). (2005). Systems for state science assessments. Washington, DC: The National Academies Press.

National Research Council (NRC). (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: The National Academies Press.

Schwarz, C.V., Reiser, B.J., Davis, E.A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632 – 654.

Songer, N.B., & Gotwals, A.W. (2006). The development and evaluation of an assessment system to chronicle urban students learning progressions associated with deep thinking in science. San Francisco, CA: Paper presented at the Annual Meeting of the. American Education Research Association (AERA).

Venville, G. J., & Treagust, D. F. (1998). Exploring conceptual change in genetics using a multidimensional interpretive framework. *Journal of Research in Science Teaching*, 35, 1031-1055.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching, 46*(6), 716 – 730.

# What Counts as Scientific Practice? A Taxonomy of Scientists' Ways of Thinking and Doing

Lori M. Takeuchi, The Joan Ganz Cooney Center at Sesame Workshop, 1 Lincoln Plaza, New York, NY 10023
lori.takeuchi@sesameworkshop.org

**Abstract:** Education reformers advocating the use of GIS in K-12 science classrooms claim that the tool provides authentic contexts for students to think and act scientifically. However few studies, if any, have empirically investigated the latter. I observed a marine ecology laboratory to determine what counts as scientific practice, and then an 8th grade oceanography classroom to see whether and how these practices played out there, especially around participants' use of GIS software. This paper focuses on the methods I used to draw meaningful comparisons between the novice and expert settings I studied. In particular, it describes the taxonomy of scientific practices derived from my laboratory observations, and then the nature of work in the lab as seen through this framework.

## Introduction

As globalization and rapid advances in technology change life as we know it on a daily basis, the ability to navigate complex information, reframe problems, evaluate evidence, argue persuasively, innovate, and adapt to ever-changing circumstances has come to comprise the standard repertoire of skills required to survive and succeed in the 21st century. Consequently, preparing students to think and act scientifically is an increasingly critical concern. National standards urge teachers to engage students in the practices of professional scientists as a means of developing these habits of mind, and suggest the use of scientists' tools to support their inquiry. A geographic information system is one such tool, and though around since the 1960s, is gaining popularity among a variety of professions. Earth and environmental scientists, fast food franchise developers, epidemiologists, and others use GIS to capture, store, analyze, and display geo-referenced data, which is information tagged by latitude and longitude coordinates. Layering different data sets atop a base map can reveal spatial relationships among the data that are not ordinarily perceptible in nature.

GIS has also been in K-12 classrooms for more than a decade. In addition to fostering students' spatial skills (NRC, 2006), these tools can help science learning better resemble science practice, heightening the much-desired element of *authenticity* in science classrooms—or so education reformers have claimed (e.g., Edelson, 1998; Gordin & Pea, 1995). To empirically test the assumption that GIS promotes authentic science learning, I observed a marine ecology laboratory first to determine what counts as scientific practice, and then an 8th grade oceanography classroom to see whether and how these practices played out there, especially around participants' use of GIS software. When students have access to scientists' tools, do they use them in ways that scientists do? While the primary interest of my larger study was inquiry learning in science classrooms (Takeuchi, 2008), this paper focuses narrowly on the methods I used to draw meaningful comparisons between the novice and expert settings. In particular, it describes the taxonomy of scientific practices derived from my observations in the laboratory, and then the nature of work in the lab as seen through this framework. As far as I am aware, no other such taxonomy exists. Limitations of the catalog and its implications for science education research are also discussed.

## Past Approaches to Studying Authentic Learning

What counts as scientific practice? Would I recognize authenticity in the 8th grade classroom when I saw it? To be sure, the field of science studies (1) has over the past quarter century generated a multitude of models of scientific practice, but none to date have examined scientists who use GIS per se. And while other studies of authenticity in science learning reference models of professional practice, few have directly compared the naturally occurring activities of novice and expert scientists. Rosebery et al (1992) and Krajcik et al (1999), for instance, focused on classroom inquiry but made comparisons with the generic, composite portraits of working scientists advanced by the science studies literature. Other studies have directly compared experts and novices, but on simulated tasks (e.g., Kuhn, 1989; Lowe, 1999), raising issues of ecological validity.

Another approach to studying the practices of novice and expert scientists involves the comparison of archival accounts of scientific work. Since this approach makes use of audio- or videotaped footage of naturally occurring activity, researchers are able to take into account the situational features that simulated tasks cannot. Hall (1999), for example, examined videotaped footage of people "having a theory" in a medical school classroom and around a family dinner table. Similarly, Koschmann and Zemel (2006) compared audio- and videotaped accounts of professional astronomers "doing discovery" at an observatory and two high school students doing the same using a physics simulation on a computer. Since comparisons were based upon other people's recordings, interpretations of recorded events were necessarily mediated through the original researchers' selection, framing, and view of events. Another limitation of this approach is that researchers have limited access to information about the participants or about the broader contextual features of each situation, as the events depicted in Hall (1999) and Koschmann and Zemel's (2006) studies had long passed by the time these authors got around to analyzing them.

My research is more closely modeled after Stevens' (1999) first-hand accounts of activity in the novice and expert settings he set out to compare by intent. He investigated everyday math activities as they emerged in a middle school design project and in the real-world design projects of an architecture firm. In the seven months he spent in the classroom and year observing architects, he witnessed how tools such as CAD software, available to both students and architects, mediated project activity in ways reflecting the culture of each institution. Comparing the forms of mathematical practices across the two settings, Stevens concluded that the organizational environment of the school—e.g., an assessment system that values equality over diversity, and the structure of the school day—makes it difficult for students to engage in meaningful mathematical practices and, consequently, learn very much.

## Methods

I adapted the design of Stevens' (1999; described above) study to investigate how scientific practices emerge around the use of GIS in an 8th grade oceanography classroom and a university marine ecology laboratory. The unit of analysis in each setting was the activity system (Engeström, 1999) of a small team conducting a research investigation. The laboratory case team consisted of a marine ecology doctoral student and his field assistant, who were studying the movement patterns of a particular fish species. The classroom case team comprised two 13-year-old boys investigating the kelp forest ecosystems of Catalina Island. I collected data intermittently over a 9-month period in the laboratory, and over 11 consecutive weeks in the classroom. In both settings I took field notes, conducted interviews, collected artifacts, and videotaped human-human and human-computer interactions.

### Operationalizing "Scientific Practice"

To identify scientific practices in the laboratory, I first had to define "scientific practice." Operationalizing the concept to support meaningful comparisons between laboratory and classroom settings proved challenging: Framing it too closely around the technical aspects of scientists' work, for example, provided too little to compare or even contrast. Crafting a definition around the inquiry processes we want students to develop, on the other hand, rendered the comparison to professional practice extraneous. In the end, I combined aspects of Rouse (1994), Pickering (1992), and Edelson's (1998) ideas into a definition that I thought would best serve the purposes of this study:

> Scientific practices are sustained, dynamic, and communal acts of purposeful inquiry. These "patterns of activity in response to a situation" (Rouse, 1994) manifest the beliefs, values, and experiences accumulated by the scientific community over time, and include ways of thinking and doing, attitudes, and social interactions.

This definition does not include scientists' tools and resources per se, and it intentionally omits formalized investigative procedures like *run experiment* or *analyze data*. I instead use these organized activities as framing contexts for their more spontaneous acts of scientific thinking, doing, feeling, and interacting. Moreover, since I observed all phases of the students' investigation but just the scientists' data collection phase, the case teams' individual sets of procedures would have been incommensurable.

### Coding for Scientific Practices

Several science studies scholars have contributed to theoretical notions of scientific practice (e.g., Latour & Woolgar, 1987; Knorr-Cetina, 1981; Pickering, 1995; Rouse, 1994). However, I was unable to locate previous work that comprehensively indexed a practical list of scientific practices that could serve as the basis for my cross-setting comparisons. I myself would have to empirically determine what counts as scientific practice. So, once data collection was complete, I used the above definition to generate codes for scientific practices seen in the laboratory. Through the process of analytic induction (Znaniecki, 1934) and deduction, each practice was assessed for its generality across the corpus of data and revised or dropped if counterexamples were found. I made several passes through the data to ensure that all practices fit the working definition—itself revised several times—and that none were subsets or duplicates of others. While the present set of codes is by no means comprehensive, it is a starting point for more systematic efforts to concretize the rather abstract notion of scientific practice.

I then used HyperRESEARCH qualitative data analysis software to code instances in the laboratory case's field notes and video content logs for the identified practices. I also chunked these data by *task*, which I've defined as "a clearly delineated piece of work that is of short or limited duration and performed to move the investigation forward." The task served to bound the scientists' work activities into discrete units of analysis so that scientific practices could be tallied within each task. However it is important to note that these tallies may be more indicative of my observational sampling than the quality or complexity of the corresponding tasks.

### The Laboratory Setting and Participants

Kinder Marine Laboratory (all names are pseudonyms) is a large, ocean-side research facility of a prominent West Coast university. Between July 2006 and April 2007, I observed marine ecology doctoral student Fred (age 33) and his field assistant Kyle (age 28) during the fourth of Fred's five-year long dissertation study. Fred's research focused on the movement patterns of individual spotted purplefish in relation to features of the local purplefish population (size, density, reproductive behavior) and environment (bottom types, biological habitat). Findings would have implications for the design of marine protected areas (2), since the effectiveness of these reserves depends on the extent to which they accommodate the full range of movements of the individuals they aim to protect. I started observing the team one year after Fred began collecting data, and just one week after Kyle started as Fred's field assistant. The case scientists belonged to the smaller, 20-person Coniglio-Ross Lab for Marine Ecology Research at Kinder (nicknamed the CR-Sea Lab), and were the only ones there at the time using GIS in their research. In fact, when I met Fred, he was just learning GIS because he knew the tools would facilitate his research.

I observed Fred and Kyle in three locations: (a) at the CR-Sea Lab; (b) *above* Fred's research site out on Cabezon Harbor (I went out on their research vessel the *Otter's Nest* but did not accompany them underwater on their scuba dives); and (c) inside the van they used to transport the *Otter's Nest* between these two locations. Although I focused my attention on Fred and Kyle's activities, I also captured some of their interactions with members of the wider Kinder community, including faculty scientists, technicians, and other graduate students.

## Scientific Practices Observed in the Laboratory

In my 40 hours observing and interviewing the case study scientists, I recorded the following 45 scientific practices. They are listed in Table 1 order of frequency logged, from highest to lowest. Since I did not visit the scientists on a daily basis, and was not able to observe them over the complete course of their five-year-long project, the practices on this list do not represent the complete repertoire of their work. As such, the corresponding frequencies should also be taken as relative indicators and a specific reflection of the days I chose to observe there. Following this list are discussions of the top 15 observed practices. (See Takeuchi, 2008 for descriptions of all 45 practices.)

Table 1. Scientific practices observed with the case study scientists

| Practice | Frequency | Nature | Practice | Frequency | Nature | Practice | Frequency | Nature |
|---|---|---|---|---|---|---|---|---|
| Use inscription | 34 | d | Connect study to personal exp. | 9 | e | Be meticulous/organized | 5 | d e |
| Maintain equipment | 20 | d e | Hypothesize | 9 | e | Communicate data/findings | 5 | d e |
| Divide labor/expertise | 18 | d | Partake of comm. resources | 9 | d | Compete for resources | 5 | d |
| Accomm. natural phenom. | 17 | d | Reason with others | 9 | d e | Discover/Achieve insight | 5 | |
| Identify pattern | 17 | d | Seek help | 9 | d e | Make meaning | 5 | d |
| Let tool shape research | 17 | d e | Seek new knowledge | 9 | d e | Put safety first | 5 | e |
| Apprentice | 15 | d | Strive for reliable data | 9 | e | Reframe problem | 5 | e |
| Follow ritual | 13 | d | Tweak tool/technique | 9 | e | Take risk | 5 | e |
| Save time/resources | 13 | d | Ask question | 8 | e | Be driven by ulterior motive | 4 | e |
| Accomm. misbehaving tool | 12 | d | Camaraderie | 8 | d | Deal with messiness of data | 4 | d e |
| Be committed/persistent | 12 | e | Modify plans/thinking | 8 | e | Ideate | 4 | e |
| Master routine | 12 | d | Participate in community | 7 | d | State limits of data/tool | 4 | e |
| Plan/Prep | 12 | d e | Appropriate new tool | 6 | e | Connect study to world issue | 3 | d e |
| Be uncertain | 10 | e | Tinker with tool | 6 | d e | Do politics | 3 | d e |
| Reason with tool | 10 | d | Apply background knowledge | 5 | d e | Invent | 3 | e |

Notes: d = Practice is a manifestation of or response to the *distributed* nature of work.
e = Practice is a manifestation of or response to the *emergent* nature of work.

### 1. Use Inscription (34)

Fred and Kyle used inscriptions (3) in nearly every task I observed, plus the ones that took place underwater. They printed waterproof maps of the harbor's depth to serve as underwater guides to the seafloor; analyzed individual fishes' home ranges on the GIS; and projected slides of fish tag expiration timelines on the lab wall to solicit help from advising professors, among other things. As the study's principal investigator, Fred spent much more time than Kyle did constructing and managing inscriptions for the planning, data collection, analysis, and communication activities they performed together.

## 2. Maintain Equipment (20)

While Fred took care of preparing inscriptions, Kyle was often occupied with maintaining the *heavy* equipment required to do their scientific work, including the 30-foot research vessel the *Otter's Nest*, the trailer that transported the *Otter's Nest* to Cabezon Harbor, and the rusty old van that towed the trailer 100 miles roundtrip to and from the harbor. Equipment maintenance involved packing up transport vehicles with equipment used on dives, filling vehicles with gas, fixing broken items, cleaning up and storing equipment, and making arrangements for regular servicing. Kyle also maintained the team's scuba tanks and the four buoys—each weighing 43 kilograms—that served as receivers for the acoustic tracking system set up in the harbor. I emphasize *heavy* above because the lighter equipment associated with inscriptional work—i.e., Fred's laptop, the lab's GIS desktop computer, and the GPS unit—for the most part fell under Fred's jurisdiction, as did the tiny and expensive acoustic tags that were surgically implanted into the fish for tracking by the telemetry system.

## 3. Divide Labor or Expertise (18)

Fred and Kyle often assumed responsibility over separate tasks in order to accomplish the maximum amount of work out of available time. Divisions of labor usually fell along lines of expertise, professional rank, location, or individual commitments. Fred, for example, managed the fish tag equipment because he possessed more specialized knowledge about the technology than Kyle did. And it was Kyle's job to load the van up each morning at 6:30 AM because this is the type of grunt work that field assistants are hired to do. On lab days—versus days spent in the field—Fred often worked inside completing paperwork, writing research grants, or working with software, while Kyle could be found outside in the storage sheds fixing or maintaining equipment. In many cases, work could only be accomplished on this two-man team if tasks were divided. Fred, for example, always steered to position the boat using his GPS unit as a guide while Kyle dropped anchor. Divisions of expertise also contributed to the efficiency of the work team, as when Fred had Kinder's resident geneticist run DNA tests on the fish eggs he collected.

## 4. Accommodate Natural Phenomenon (17)

As much as scientists attempt to maintain control over all aspects of their research, nature often evades management. Fred and Kyle had to adjust their plans, tweak their equipment, and redesign their investigation to accommodate the occasional unforeseen and uncontrollable behaviors of the fishes under study. In the spring, when Kyle was having no luck locating egg nests of the fishes they had tagged the previous summer, Fred decided to extend their egg hunts beyond his study site to untagged fish. Doing so would increase his sample size, but disassociate a set of genetics tests he wanted to run from the home range aspect of his study. The scientists' work was also at the mercy of the sea and the sky. Environmental events—e.g., high surf, cold water temperatures, algal blooms—regularly altered their course of action on a moment-by-moment basis.

## 5. Identify Pattern (17)

Scientists often craft visualizations of their data to reveal patterns that remain hidden in rawer formats. These patterns help them make predictions and make sense of the represented phenomena. In a laboratory meeting convened around the mysterious disappearance of several fish Fred had tagged earlier in the year, the participating scientists rarely referred to one of the displayed graphs, tables, or maps without extracting some less-than-obvious pattern from it. The three senior scientists in the room surfaced patterns in Fred's data that even Fred had never seen. They were so generative in this regard, in fact, that it was as though they couldn't look at an inscription without looking for underlying patterns. By meeting's end, their interpretations of the data led the group to the conclusion that Fred had perhaps ill advisedly constrained his study site to the published receiving range of the buoys, and needed to expand it to cover the region that his tagged fish actually inhabit.

## 6. Let Tool Shape Research (17)

The scientists often allowed the affordances and constraints of their tools influence how they designed or proceeded with their investigation. What Fred and Kyle managed to accomplish in a day's data collection expedition, for example, depended on the number of scuba tanks they brought with them on the boat (also limited by the size of the boat) and the oxygen composition of each tank. A tank holding 1,200 pounds of air gave each diver 45 minutes to complete a fish density or UPC survey of one reef in their study site. The oxygen-enriched tanks with a 20/80 oxygen to nitrogen mix allowed Fred and Kyle to perform more physically exerting tasks, such as underwater fishing (4), in the same amount of time, but were harder to obtain at Kinder. Sometimes Fred and Kyle's air supplies dictated their dive itineraries. As Fred once instructed Kyle just before their final dive of the day, "Let's just make sure that we come up with enough air to switch the buoy out, and if we catch one (a fish), that we come up with enough air to bring it back down."

7. Apprentice (15)

Fred hired Kyle as his field assistant just two weeks before I started observing them, so I got to witness a critical stretch of Kyle's on-the-job training. Very little time was set aside for direct instruction. Instead Kyle was expected to watch Fred perform a procedure once through, such as locating and unmooring the telemetry buoys, or preparing their rods and reels for underwater fishing, before taking responsibility for the task the next time. Although it was up to Kyle to direct his own learning—by asking Fred questions rather than waiting for Fred to provide lock-step instructions—Fred was fully cognizant of the training session in progress. In fact, when there were new techniques for Kyle to learn, Fred often made sure his apprentice stayed near enough to see him model the activity. These learning episodes were among the rare occasions that the teammates worked on the same task simultaneously.

8. Follow Ritual (13)

During the summer field season spanning July through October, the scientists ran scuba expeditions four days of the week to collect data and maintain their telemetry equipment on the harbor. Field days began at 6:30 AM loading up the boat and van at the lab, and included a 100-mile round trip drive to and from Cabezon Harbor, where they took three or four scuba dives over the course of four or five hours. Days often ended past 6:00 PM, after hosing down and storing all of the equipment. Between tasks, scientists participated in a number of social rituals to take the edge off of this grueling work. Morning trips to Emmy's Bakery for muffins and coffee, for example, were mandatory even when the scientists were running behind their on-the-road-by-7:00 schedule. Other social rituals included the "hot water bucket" that researchers would pour down their wetsuits after their last dive of the day to warm their bones, and chitchat on the long drives between Kinder and Cabezon Harbor. These commonly shared experiences helped build rapport within the Kinder community, and were passed down from old-timers to newcomers.

9. Save Time or Resources (13)

Fred and Kyle constantly sought to make the most of their limited time and resources. Some of these efforts were built into routines—as described in *divide labor* and *apprentice*, which are not included in this count—but others involved a conscious pursuit of the most efficient path through their work. Fred, for instance, spent a couple of days constructing and then printing 25 percent contours of the tagged fishes' home ranges so that on scuba dives they could more easily locate egg nests to collect samples. More time spent at the computer meant less time swimming around underwater in search of the nests, expending energy and oxygen stores. Kyle's efficiency-motivated acts include creating drop-down menus in an *Excel* file to speed up data entry and minimize error, and inventing a special belt holding eight large vials and cuticle scissors to make it easier to harvest fragile fish eggs while underwater.

10. Accommodate Misbehaving Tool (12)

Just as the scientists altered their plans or even goals in response to the unanticipated behaviors of the fish, the weather, or the sea, they also accommodated the equipment that broke down or failed to produce some desired outcome (Pickering, 1995). Small misbehaviors, such a leaky O-ring on a scuba tank, might alter the day's data collection agenda, but bigger ones affected changes at a greater scale. The disappearance of half of Fred's tagged fish from telemetry reception, for example, prompted him to spend several weeks catching more fish to tag and include in his study. It also prompted him to rescale his study site in order to monitor a wider region of Cabezon Harbor and, hopefully, to recover reception of those missing fish. Fred and Kyle eventually solved the mystery of the disappearing fish—the batteries in their tags had simply died—but the time, effort, and expense spent in making these adaptations were significant. In a final act of accommodating the faulty fish tags, Fred decided to set the end point of the data collection phase to whenever the batteries in his remaining tagged fish petered out.

11. Be Committed or Persistent (12)

Only the most devoted make it in this field. Based on physical demands alone, those who can't take five months of 12-hour field days and scuba diving in 54-degree waters are weeded out at the level of field assistant. Fred's acoustic telemetry system required that he cycle a charged buoy into the harbor every week of the year, including stormy winter months, so that it would continually monitor Fred's tagged fish without interruption. The considerable commitment of these scientists extended beyond physical work. The unpredictable and often unwelcome behaviors of their fish and tools *(accommodate natural phenomenon* and *misbehaving tool*) often required extra time, money, and exertion to put the study back on track. Patience and perseverance are necessary to deal with the uncertain outcomes and regular disappointments associated with this type of work (Edelson, 1998; Kurz-Milcke et al, 2004):

> Fred: All these tagged fish that I tagged last summer... I did it for two years, right, and they all stuck around for the entire two years. And so I did them all last summer thinking, "Oh, this summer I'll just jump right into manipulating all this stuff." And now I just came back and half of

> them are gone! Ahhhh! It's really disappointing. I mean, but that's part of it. And you get other data from it. But I can totally see like if you don't have that interest and commitment to it, you can get that stop-doing it attitude. It's definitely kind of long and ongoing and always changing.

### 12. Master Routine (12)

The scientists repeated many of the same routines—defined here as regularly repeated procedures—on a weekly, if not daily, basis. Weekly routines included the buoy switching mentioned above. This involved unmooring the least-charged of three telemetric buoys from the harbor floor and transporting it back to the lab for recharging and replacement the following week. Daily routines included loading up the van and boat with the necessary dive and data collection equipment each morning, and unloading it at the end of the day. Certain routines involved complicated procedures that both Fred and Kyle became more skilled at with practice, as they did with their underwater fishing techniques. Other routines, such as loading and unloading transport vehicles at the lab, were familiar practices among members of the wider Kinder community. On the day that Fred couldn't dive due to an ear infection, the field assistant he "borrowed" from another crew was able to step in and perform a number of these institutionalized routines with little instruction.

### 13. Plan/Prep (12)

Due to the tremendous material aspect of their work, along with the fact that it often took the coordinated actions of several people to accomplish a single task, the scientists spent quite a bit of effort orchestrating upcoming tasks. Poring over the minutiae of their data collection, analyses, and presentation activities beforehand helped Fred and Kyle obviate unwelcome complications. It also ensured that minimal time and money were wasted in the implementation. The scientists often spent far more time prepping for than executing target tasks, as examples cited in *save time or resources* suggest. Fred spent months preparing his dissertation proposal, which, once approved by his dissertation committee, served as the official blueprint for his study. But most of the planning I saw took place on a more impromptu basis—on van drives to and from the harbor, while suiting up in the parking lot, or just before putting their face masks before a dive. Fred and Kyle were in constant communication about how to best coordinate their upcoming actions.

### 14. Be Uncertain (10)

Never knowing what to expect next in an unfolding investigation—especially ones marked by *misbehaving tools* and *natural phenomena*—instills in scientists a comfort with uncertainty. The scientists demonstrated this comfort by openly admitting to colleagues what they didn't know, couldn't understand, or were willing to accept as unanswerable, in one-on-one and lab-wide meetings alike. They also demonstrated this comfort in their ability to carry on in their work and make important decisions despite unknown outcomes. Kyle, for example, spent several hours fabricating a special belt to facilitate their underwater egg sampling excursions using random materials he pilfered from around the lab. He had no idea if the belt would work—he had yet to try it out at the time of our interview—since the whole egg collection process was uncharted territory for Fred and Kyle. A common corollary to the scientists' uncertainty was a desire to bring clarity to a situation, and this required a high tolerance for risk. As Fred's advisor Joe pointed out while counseling Fred on whether to shift the scale of his study midway through the investigation, "It's just one of those things that you have to take a shot at and see what's going to happen. That's the way science is. It sucks! Ultimately [...] you're trying to figure out what's the best effort to put my time into."

### 15. Reason with Tool (10)

The scientists reasoned with tools when they used artifacts in their environment to (a) offload complex cognitive tasks or (b) extend their cognitive capabilities. In an example of the former, Fred used a dive watch to help monitor his oxygen usage so that he could instead focus on his underwater tasks. In an example of the latter, he calculated statistics using software. Fred reasoned with the GIS for both purposes. Here Fred describes the advantages the tool brought to his research:

> Well, one thing that it really does the most for me is just to think about all this stuff. To just sit in front of it and actually look at it instead of just having a number. And you can get these distributions of points where the fish is and then you calculate an area and then you know the home range is 500 square meters. But actually looking at the shape and seeing how that fits into the habitat, really that's a big help, I think. And kind of coming up with ideas, and thinking about, "Oh this looks interesting, I should actually do the stats on that."

The spatial and layering capabilities of the GIS freed Fred up from having to translate the text data into something meaningful, and inspired analyses that he wouldn't have come upon unaided by the spatial display.

## The Distributed and Emergent Nature of Science

In the process of sorting through the 45 observed practices, I characterized the science that took place at Kinder Marine Laboratory to be both *distributed* and *emergent* (see tags in Table 1, above). I also established a pattern of *efficient* and *adaptive* responses to the problems presented by distributed and emergent systems. Though these findings merely confirm what empirical research in science and technology studies documented in the 1980s and 1990s, they are worth discussing briefly here, as these constructs helped guide my analyses of the 8th-grade classroom case.

Science is no longer construed as the solitary endeavor of brilliant individuals such as Isaac Newton or Albert Einstein. Scientific research and progress is more readily recognized as the joint achievement of people, organizations, places, and tools, all of which are distributed across space and time (Goodwin, 1994; Dunbar, 2000). I observed Fred and Kyle participate in 28 practices that were manifestations of and/or responses to the distributed quality of their work. Twelve such practices describe social interactions (e.g., *divide labor/expertise, apprentice, partake of community resources, seek help, participate in community, communicate data/findings, compete for resources, reason with others, do politics*), and seven describe their relationships with artifacts (i.e., *use inscription, maintain equipment, let tools shape research, identify patterns, reason with tool, accommodate misbehaving tool*, and *tinker with tool*). This dispersion of knowledge, reasoning, and labor across individuals, artifacts, institutions, and spaces both facilitate and pose particular challenges to the accomplishment of work.

Fred and Kyle often responded to the challenges with efficient measures. Some of these efforts were built into their routines, such as dividing tasks and Kyle shadowing Fred in an apprentice style of learning. But others involved a conscious pursuit of the most efficient path through their work. In some of the examples cited above—e.g., mapping out dives on the GIS to spend less time underwater, or creating drop-down menus in Excel to facilitate data entry—the scientists "sacrifice short-term efficiency to retool [their] knowledge or context for the prospect of long-term gain," in what Martin and Schwartz call a "prospective adaptation" (in press, p. 5). Because Fred's investigation took five years from start to finish, the time investment on any prospective adaptation paid itself off many times over.

Efficiency in the form of procedural speed and accuracy benefits distributed work systems, especially when these systems are stable and see little change. But the work of Fred, Kyle, and most scientists are far from predictable. No matter how meticulously they plan out their investigations, circumstances beyond their control have a way of altering the science that gets done. Pickering (1995) attributes the "temporally emergent" structure of scientific research to the nonhuman or material agency of scientific practice. When a machine fails to produce expected outcomes, humans often accommodate the machine by revising their original goals, or by tweaking the machine or the investigation itself. Other researchers (e.g., Knorr-Cetina, 1981) have documented how the particularities of a research situation—material, personal, or otherwise—influence discoveries in unexpected ways. What becomes scientific fact depends, to a great extent, on the context in which the science is conducted.

This emergent quality of science is also revealed in the taxonomy, with 30 of Fred and Kyle's observed practices falling into this category. I observed the scientists adapt to unanticipated events (i.e., *let tools shape research, accommodate natural phenomenon, accommodate misbehaving tool, appropriate new tool*, and *tinker with tool*), and take action to either anticipate the unknown (e.g., *maintain equipment, plan, seek new knowledge, strive for reliable data, hypothesize, take risk*) or rearrange the environment to handle new problems or information (e.g., *seek help, reframe problem, ideate, invent*). The emergent nature of their work also elicited certain attitudes: *commitment* is what kept Fred and Kyle coming back to work day after day in the face of recurrent surprise and disappointment, and their comfort with *uncertainty* is what permitted them to do so.

## What else counts as scientific practice?

Once developed, I took this taxonomy into an oceanography classroom, where it guided my analyses of a pair of 8th graders' GIS-based investigation of Catalina Island's kelp forest habitats. The students participated in many of the same practices I observed at Kinder Marine Laboratory, suggesting that the GIS provided an authentic context for their science learning. However, they were less inclined to engage in practices classified as adaptive than the scientists were, with evidence indicating that aspects of the school setting steered students toward more efficiency-oriented patterns of response. For a more detailed discussion of the classroom case study vis-à-vis this taxonomy, see Takeuchi, 2008.

Although the 45-practice taxonomy was sufficient for the types of comparisons I drew between the laboratory and classroom settings, it represents just one year (elapsed) of the scientists' work and primarily the data collection phase of their investigation. I imagine that others in the science studies and science education communities might also find value in such a catalog, but to be useful to a wider audience, the list would need to capture the complete investigation cycle. I would enjoy returning to the field to expand and refine the current list by watching practicing scientists over a longer time frame, and by augmenting the data on marine scientists with observations of other types of natural scientists (e.g., seismologists, meteorologists, stem-cell researchers, astronomers). The resulting universal catalog of scientific practices would provide a powerful set of images that could be used for practical and theoretical purposes in science classrooms, research, and policymaking.

## Endnotes

(1) Science studies is a blanket term for the genre of research that examines what scientists do and the nature of scientific knowledge. Within this genre there are a number of disciplines with their own specific methodologies and epistemologies (e.g., sociological studies of knowledge, philosophy of science, postmodernism).

(2) A marine protected area (MPA) is any area of the marine environment that has been reserved by law (federal, state, or local) to protect endangered plant or animal species, ecosystems, and important historical and cultural sites.

(3) Scientists' representations of their objects/phenomena under study. Inscriptions are materially embodied in some medium, such as paper or computer monitors, and include maps, graphs, photographs, equations, diagrams, and hand-scribbled notes.

(4) To implant the fish with acoustic tags, Fred and Kyle had to gently capture them from the seafloor and bring them up to the boat's deck for surgery. They took 2-foot long fishing rods underwater with them and dangled hooks baited with live shrimp right in front of individuals of the target species.

## References

Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology, 21*(1), 49-58.

Edelson, D. C. (1998). Realising authentic science learning through the adaptation of scientific practice. In K. G. Tobin & B. J. Fraser (Eds.), *International handbook of science education* (pp. 317-332). Dordrecht, Netherlands: Kluwer Academic Publishers.

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engestrom, R. Miettinen & R.-L. Punamaki (Eds.), *Perspectives in activity theory* (pp. 19-38). New York: Cambridge University Press.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606-633.

Gordin, D. N., & Pea, R. D. (1995). Prospects for scientific visualization as an educational technology. *Journal of the Learning Sciences, 4*(3), 249-279.

Hall, R. (1999). The organization and development of discursive practices for "having a theory." *Discourse Processes, 27*(2), 187-218.

Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal of Research in Science Teaching, 24*(4), 291-307.

Knorr-Cetina, K. D. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press.

Koschmann, T., & Zemel, A. (2006). *Optical pulsars and black arrows: Discovery's work in "hot" and "cold" science.* Paper presented at ICLS 2006, Bloomington, IN.

Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., & Fredricks, J. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences, 7*(3&4), 313-350.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674-689.

Latour, B., & Woolgar, S. (1986). *Laboratory life*. Princeton, NJ: Princeton University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lowe, R. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education, 14*(2), 225-244.

Martin, L., & Schwartz, D. (in press). Prospective adaptation in the use of external representations.

National Research Council. (2006). *Learning to think spatially: GIS as a support system in the K-12 curriculum*. Washington, DC: National Academy Press.

Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago: The University of Chicago Press.

Rosebery, A. S., Warren, B., & Conant, F. R. (1992). Appropriating scientific discourse: Findings from language minority classrooms. *Journal of the Learning Sciences, 2*(1), 61-64.

Rouse, J. (1994). Engaging science through cultural studies. In *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. Two: Symposia and Invited Papers, pp. 396-401).

Stevens, R. R. (1999). *Disciplined perception: Comparing the development of embodied mathematical practices in school and at work.* Unpublished Dissertation, University of California at Berkeley, Berkeley, CA.

Takeuchi, (2008). *Toward authentic scientific practice: Comparing the use of GIS in the classroom and laboratory.* Unpublished Dissertation, Stanford University, Stanford, CA.

Znaniecki, F. (1934). *The method of sociology*. New York: Farrar & Rinehart.

# Analyzing Collaborative Knowledge Construction in Secondary School Biology

Vanessa L. Peters, James D. Slotta, Ontario Institute for Studies in Education, University of Toronto
252 Bloor Street West, Toronto, Ontario, Canada M5S 1V6
Email: vlpeters@gmail.com, jslotta@gmail.com

**Abstract:** This research investigates student collaboration in a high school biology curriculum that was based on the Knowledge Community and Inquiry (KCI) model. Using co-design, the researchers collaborated with three high school science teachers to design a curriculum where 112 grade-ten biology students collaboratively developed a community wiki about Canadian ecozones and biodiversity issues. Students then used the wiki as the primary resource for a subsequent inquiry activity. This paper analyzes students' contributions to the knowledge base, test performance, and student satisfaction to evaluate the efficacy of the KCI model. A new method of analysis for collaborative wiki artifacts was developed to measure student interaction in the wiki. We found that students who were higher contributors to the knowledge base performed better on a post-test than students who were measured as low contributors. Our findings suggest that the KCI model is a promising mechanism for supporting both collaborative and individual learning.

## Introduction

Increasingly, scholars are emphasizing the importance of preparing students for the complex problems of a "knowledge society" (Drucker, 1959; Bereiter, 2002). Such 21st century skills are important for students to become productive citizens in a technological and knowledge-oriented society (Lemke, Coughlin, & Reifsneider, 2009). Such skills cannot be taught in isolation; rather, they must be learned in the context of a knowledge community where individuals collectively pursue the advancement of ideas. Many educators have noted the mismatch between conventional school curriculum and the needs of citizens in a knowledge society (Tapscott & Williams, 2006; diSessa, 2000). While schools struggle to raise student achievement on basic skills, scholars have identified a full spectrum of cognitive and metacognitive skills that are not being addressed, including inquiry, critical thinking, design and collaboration (Slotta & Linn, 2009). Thus, the traditional forms of lecture and homework that still predominate in today's classrooms are not desirable models for the design of technology-enhanced learning (Collins & Halverson, 2009).

The domains of science and technology are particularly reliant on high levels of knowledge work, yet in classroom settings teachers routinely employ rote learning in an effort to cover the mandated curriculum content (Slotta & Linn, 2009). This is particularly the case in secondary school science, which covers more topics than any other subject, thus resulting in textbooks that have been described as being "a mile wide and an inch deep" (Schmidt, McKnight, & Raizen, 1997, p. 62). Secondary teachers are responsible for teaching specific content matter, making it difficult to design learning activities where students can pursue a deep understanding of science topics. All science lessons must fit within a tight schedule of content coverage, with outcomes that are assessable by conventional measures. The fast pace of most science courses leaves teachers little or no time to ask their students "big-picture" questions, or engage them in personally relevant topics.

This paper begins by considering the established research traditions of knowledge communities and scaffolded inquiry, as well as a recent pedagogical model that describes how these two traditions can be blended to create engaging new curriculum for secondary science students (see Slotta, 2007; Slotta & Peters, 2008). We discuss the important innovation of co-design (Roschelle, Penuel, & Sechtman, 2006), which offers a means for creating such curriculum in a way that meets teachers' expectations without undermining the researchers' objectives. We discuss an eight-week design study where a technology-enhanced curriculum was co-designed according to the Knowledge Community and Inquiry model. In our analysis, we evaluate students' contributions to the community knowledge base, the effects of such contributions on test performance, and student satisfaction with the curriculum unit.

## Theoretical Framework

A number of promising instructional approaches have already shown to engage students in deep collaborative inquiry. In the research program called Fostering a Communities of Learners (FCL), Brown and Campione (1996) scaffolded student learning within a peer community. The key components of FCL – research, information sharing (jigsaw), and consequential tasks (performing) – provide structure and support for students' collaborations within the learning community. When combined, these components form an effective instructional strategy that fosters critical reflection and deep understanding of disciplinary content. An important characteristic of FCL activities is that they are independently purposeful, yet they cohere to form a functional system. Although FCL is scripted, it must be emphasized that simply following a series of steps is not enough – teachers must understand the goals and philosophies that underlie the approach.

An important theoretical contribution of FCL is the notion of *diverse expertise*. In their classroom implementations of FCL, Brown and Campione (1996) found that students came to highly value the contributions of their peers. These contributions were not always about content, they were often related to using the computers or managing the group (Collins, Joseph, & Bielaczyc, 2004). This observation solidified the importance of information sharing for fostering a sense of community among learners. The students also offered suggestions for refining the design of FCL (Bielaczyc, 2006), such as recommending that information sharing occur sooner in the process rather than at the very end of the project. This suggestion resulted in an innovation called *crosstalk*, which involves students presenting their preliminary findings to the entire class. Crosstalk became a mechanism for a peer-review of ideas that would often result in students pursuing a new line of inquiry (Bielaczyc, 2006).

Another important consideration in the research literature is that of scaffolded inquiry. Researchers have developed a number of prominent pedagogical approaches that provide students with rich collaborative inquiry activities, many of which include technology-enhanced tools and materials (e.g., Linn & Hsi, 2000; Slotta, 2004; Songer, 2006). Broadly speaking, inquiry learning is an instructional approach where students research and investigate some phenomenon before making inferences about it (Kuhn, Black, Keselman, & Kaplan, 2000). For some educational researchers, the goal of inquiry learning is to teach students to work in ways that are similar to the way real scientists work. Students pose questions, formulate hypotheses, and then design investigations to test those hypotheses (White & Frederiksen, 1998). For other researchers, the experimentation cycle is just one of many possible inquiry patterns (Linn & Eylon, 2006). In all models of inquiry, the teacher does not assume the role of "sage on the stage" (King, 1993), but that of a facilitator who guides students in leading their own investigations (de Jong & van Joolingen, 1998). Students engaged in inquiry activities learn domain-specific content while developing their reasoning skills (Hmelo-Silver, Duncan, & Chinn, 2007) and their ability to understand and apply scientific principles and concepts (Schwab, 1962).

Both inquiry learning and knowledge communities emphasize the importance of collaborative knowledge construction. Here, collaboration refers to the coordinated efforts resulting from "a continued attempt to construct and maintain a shared conception of a problem" (Roschelle & Teasley, 1995, p. 70). Fischer, Bruhn, Gräsel, & Mandl (2002) describe four characteristics of collaborative knowledge construction: (a) individuals bring their prior knowledge to the learning situation, (b) there is a cause for a learning partner to share that knowledge, (c) a consensus is reached about the content or given facts, and (d) individual perspectives are integrated into a common understanding of the task or problem. This form of learning, in which students become productive knowledge workers, citizens and lifelong learners (Stahl, 2000), always occurs in a social context, either face-to-face or computer mediated. The essential element is the provision of a shared space for individuals to engage in social negotiation.

The two research traditions of scaffolded inquiry and knowledge communities provide mechanisms for deeply engaging students in science learning. However, despite widespread enthusiasm, these approaches have yet to make strong headway in science classrooms, as researchers have yet to determine how they can promote new cultures of learning while remaining sensitive to curriculum standards: What types of pedagogical and technological innovations are required to transform classrooms into learning communities? How can these innovations be designed? What supports do teachers need to enact new approaches in a manner that does not undermine the theoretical commitments of the design? These are questions that need to be addressed before inquiry-oriented instruction can pervade secondary school science curricula.

## Knowledge Community and Inquiry

In an effort to make headway on these problems, Slotta (2007) has developed the Knowledge Community and Inquiry (KCI) model which combines collaborative knowledge construction with scripted inquiry activities to target specific curriculum learning objectives (Figure 1). The model begins with a collaborative knowledge construction activity where students explore and investigate their own ideas as a community of learners, creating knowledge artifacts that are aggregated into a communal knowledge base. An important component of collaborative knowledge construction is that learning activities (such as inquiry-type investigations) must be guided by the community itself through the knowledge construction process (Scardamalia & Bereiter, 1996). Common themes, ideas or interests should emerge, reflecting the "voice" of the community. The instructor must listen to this voice and respond by designing activities that reflect students' interests. The latter process is critical, but also challenging to execute, since the design of any activity must also address the content expectations and learning goals of the curriculum.

It is no easy task to design activities that respond to community interests while addressing learning goals and adhering to time constraints. In the KCI model, scripted activities are co-designed by teachers and researchers only after the knowledge construction phase is complete, resulting in new activities that build upon the themes that were identified within the knowledge base. Students then work independently or collaboratively on the scripted activities, drawing on knowledge elements from the community knowledge base, producing new contributions to that knowledge base, and completing inquiry tasks that are directly connected to assessable content learning outcomes. KCI attempts to support the development of such curriculum by working closely with teachers to develop a carefully controlled flow of collaborative knowledge construction and scaffolded inquiry activities that are specifically designed to address the mandated curriculum.

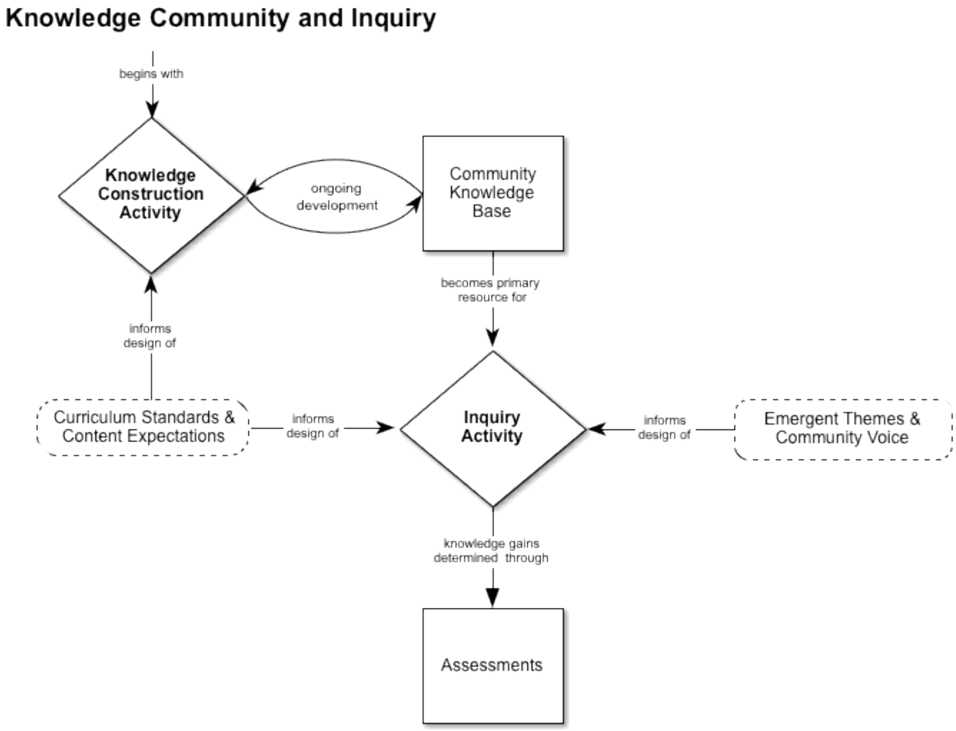



Figure 1. Flow diagram of a Knowledge Community and Inquiry curriculum.

## Methodology

The research discussed here was the second iteration of a design-oriented study that implemented the KCI model. Design research (Brown, 1992; Collins, 1992) was an appropriate methodology for this study as it supports multiple investigations across a number of contexts (McCanliss, Kalchman, & Bryant, 2003). Design research can be characterized by its iterative design cycles and formative research in authentic settings such as a real classroom (Edelson, 2002). Another important method used in this study was co-design (Roschelle, Penuel, & Shechtman, 2006), where teachers and researchers work in close partnership when designing all curriculum materials and activities. Since it was important that our biodiversity unit target the curriculum content expectations, the teachers' input was essential for creating relevant and ecologically valid materials.

### Participants and Data Sources

Participants included 112 grade ten biology students and three experienced science teachers. The co-educational school involved in this study provides specialized curriculum for high-achieving students in grades 7 through 12. Initially created as a laboratory school, this unique institution prides itself on new and innovative classroom practices. Admission to the school is competitive and based on students' score on the Secondary School Admission Test (SSAT), with 98% of new admissions being accepted from grade 6 students. The school population is ethnically diverse, with the majority of students coming from middle to upper-middle class homes. The school has a strong commitment to the liberal arts and sciences curriculum, and students are expected to fully engage in their academic program. There is a strong emphasis on community, and individual acceleration and early course specialization are discouraged. Assessment is ongoing throughout the school year and consists of formal progress reports and performance improvement plans.

### Design and Procedure

The teachers began the Biodiversity lesson by placing students into one of eight Canadian biome groups. Working in these groups, students could choose a geographical region from Canada for which they would create a wiki "Ecozone Page". A template was used to scaffold students to include specific biology content and to help them set up their wiki page. While it was important to preserve the open-ended feeling of collaborative editing that typifies wikis, it was equally important to our project that we provide students with a simple, structured way to create wiki pages that scaffolded their treatment of science concepts. We therefore created a new hybrid wiki environment that improved control over student accounts, groups, editing permissions and other features. This environment includes a special "New Page" web form (developed using Ruby on Rails) that collects metadata (e.g., with check boxes and text fields), and then generates a new wiki page with pre-specified headers to help scaffold students' wiki entries as well as the specified authoring and access permissions (see Figure 2).

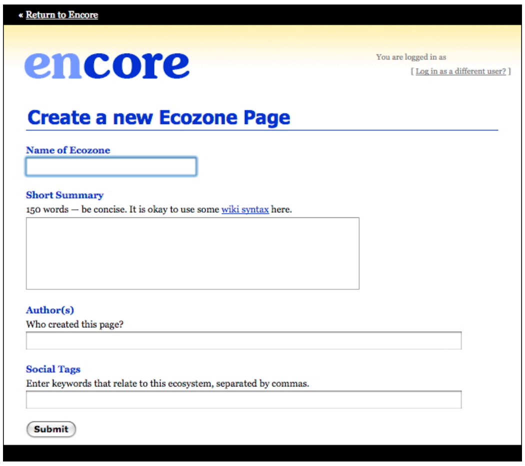



Figure 2. "New Ecozone Page" script.

Students in all four classes contributed to this same wiki repository, adding to and editing their peers' ecozone pages. After the knowledge construction activity, students were tested on their knowledge about Canadian biomes and ecozones. After working on the ecozone pages, students worked in pairs to create another wiki article about an issue or problem that posed an environmental threat to a Canadian region. Students could draw on their expertise by writing about an issue from the same ecozone they had written about in the first activity. Because biodiversity issues are not restricted to one ecozone, and to encourage synthesis of the wider knowledge base, students were asked to make connections between the different regions. Students were also asked to link to their peers' wiki pages wherever possible when referencing them in their biodiversity issue page.

## Data Sources and Analysis

Data for this study were drawn from the revision logs of students' contribution to the wiki resource, pre and post-test scores and student interviews. Content analysis was performed on a subset of students' wiki edits to gain insight into their authoring practices, and to determine the level of collaboration that was occurring. The pre and post-tests were used to gauge the effects of wiki authoring on students' test scores. Interviews revealed students' satisfaction levels with the wiki curriculum, including their perspectives of the co-authoring process. The use of multiple methods when studying collaboration is desirable since it provides a more complete picture of student engagement and interaction (Hmelo-Silver, 2003; Naidu & Järvelä, 2006).

Since there are no accepted guidelines for analyzing wiki contributions, we developed a method of content analysis to capture students' knowledge contributions to the wiki. To begin, we defined a unit of analysis for studying the wiki revision logs. The wiki technology employed in this study labels all contributions to the wiki as "changes"; however, this characterization is imprecise as it suggests that only a single change is occurring. We instead labeled such changes as individual "transactions", since it allows for multiple codes to be assigned to a particular section of the wiki (e.g., adding text, deleting text or moving text). Figure 3 is a screen shot of a revision history comparing two versions of a wiki page, and shows why this kind of segmentation (which is labeled as a single change by the wiki software), is inadequate for analysis purposes. Here, we can see two transactions: T1 and T2 (labeled as such on Figure 3). The upper portion of T1 (red text) shows a section of text as it appeared before the revision, and how it appeared afterwards (green text). The student who completed this transaction deleted some text from a paragraph (highlighted in dark red with strikethrough), and added new text (highlighted in dark green and bolded). In T2, the student added new text without altering any existing text. Referring to the content in this way enabled us to code wiki interactions without being overly reductionist in terms of the student's editing process, and permits a finer-grain analysis of wiki content.

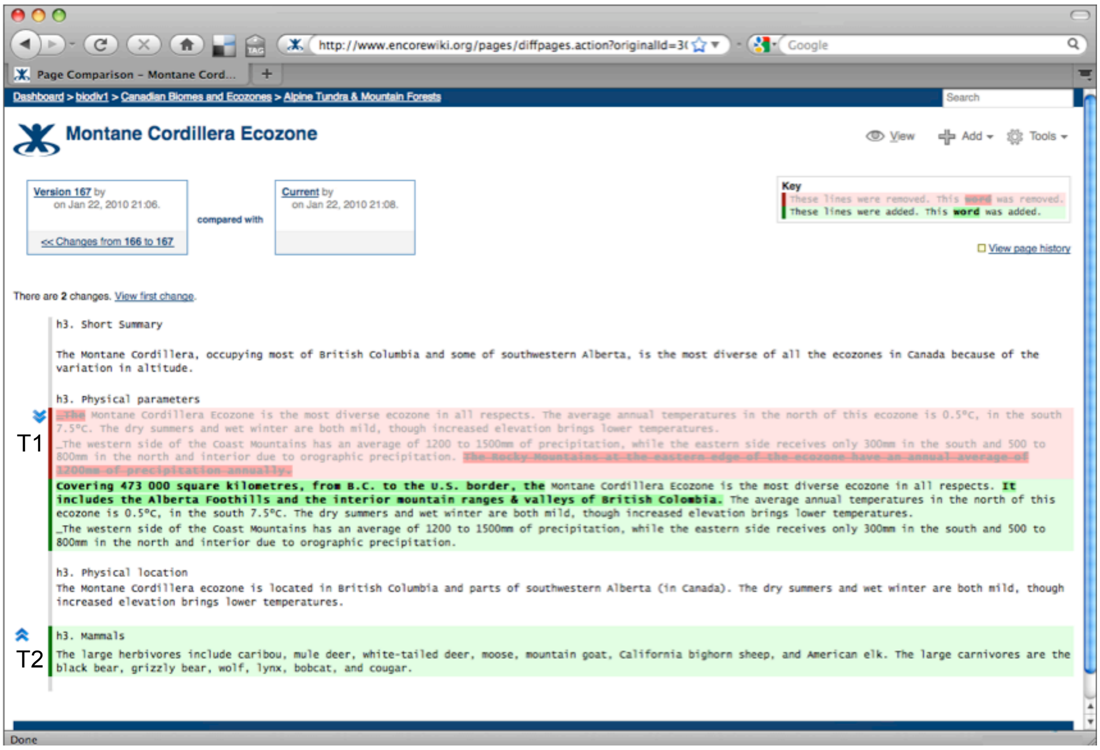



Figure 3. Screen shot of a revision history comparing two versions of a wiki page. In this comparison there are two "transactions". The first transaction (T1) contains text that was deleted and added into existing text. The second transaction (T2) contains newly added text.

We then developed a coding scheme to analyze the wiki contributions of a subset of students, coding all the transactions they made to every version of every wiki page. For the coding procedure, we first coded for "transaction type" (move, add, delete or format), then for "content type" (text, image, internal link or external link). Since substantive contributions would likely be written in prose, only text-based transactions were coded further – as either belonging to peer or self. If the edit was made to a classmate's text it was coded as peer, if it was made to one's own text it was coded as self.

## Results

Students were actively engaged in the construction of the knowledge resource within the wiki, creating 31 ecozone and biome pages, and 47 biodiversity issue pages that showed an average contribution of more than 1500 words per student ($M$ = 1651.55, $SD$ = 1700.40). Prior to analysis, each wiki page was run through Copyscape©, a web-based utility that compares web documents to check for instances of plagiarism. We found no cases that warranted concern. In the following sections we present our findings for performance gains, the collaborative knowledge construction activity and student satisfaction.

### Knowledge Construction Activity

To analyze how students were collaborating in the wiki, we applied our coding rubric to a subsample of students ($n$ = 8) to learn about their editing practices when co-authoring their knowledge artifacts. Overall, the students contributed an average of 38 transactions ($SD$ = 20.26), with high variability levels between students. Of these transactions, an average of 4.62 were images, 1.25 were internal links and 3.37 were external links. The total number of text-based transactions was 269 ($M$ = 33.63, $SD$ = 12.93)

We were interested in finding out how many of these text transactions involved *other* students text. That is, how often were students editing their own text, opposed to that of their peers? Of the students surveyed, the total number of self-based text edits was 151 ($M$ = 18.78, $SD$ = 8.87) and the total number of peer-based text edits was 118 ($M$ = 14.75, $SD$ = 9.44). The overall percentage of self vs. peer edits was roughly equal: 56.13% and 43.87%, respectively (see Table 1).

Table 1: Total text-based transactions per student: self and peer.

| | Total text edits | | |
|---|---|---|---|
| **Student** | **Self** | **Peer** | **Total** |
| Nick | 21 | 15 | 36 |
| Nathan | 22 | 31 | 53 |
| Patricia | 27 | 12 | 39 |
| Marion | 34 | 9 | 43 |
| Dave | 7 | 18 | 25 |
| Karen | 14 | 23 | 37 |
| Wendy | 15 | 10 | 25 |
| Brian | 11 | 0 | 11 |

## Performance Gains

We were interested in the effect of the knowledge construction activity on student test performance. We used only the test scores of two classes (*n* = 50) that were taught by the same teacher. We then placed these students into one of two groups based on their total number of text edits: "low contributors" and "high contributors". Students in the low contributing group were below the median for total words edited, and students in the high contributing group were above.

To find out if wiki contribution levels had an effect on test performance we conducted pre and post-tests. The difference in pre-test scores between the low and high contributing groups were not statistically significant ($t(48) = 0.2429$, $p = 0.809$), as shown in Figure 3. We did, however, find a significant difference between the two groups on a post-test ($t(50) = 4.68$, $p = <.0001$). Although this finding does not provide unequivocal evidence of higher learning gains for students who contribute more to the knowledge base, it does suggest that students who are more involved with the knowledge construction process are also engaging more with the subject content, resulting in higher test scores.

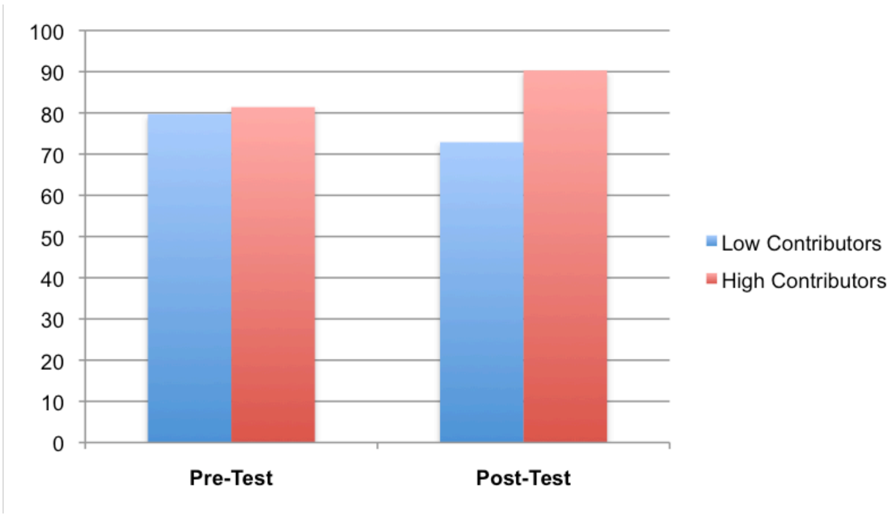



Figure 3. Test Scores from Low and High Contributing Students.

## Student Satisfaction

Student interviews were conducted in the beginning, mid, and end-point of the study. A total of twelve students participated in the interviews. Participants were interviewed in pairs in an informal environment, usually during the lunch break or after school. Since the students had never used a wiki before, we were interested in their perspectives of creating a wiki in their science class. All students responded that building a knowledge base was a fun and novel way to learn biology. One student, Marge, described her perspective of the unit as follows:

> I really enjoyed it because, like, I've never built a wiki site before and it was really cool because we actually got to make a webpage in a way – and I think it was a really interesting way of learning about Canadian biodiversity without just going on other peoples sites and memorizing all the information. Because here, I guess when you actually put the effort in

> creating a website you learn a lot more about it, or the subject itself, then when you just read it in a book – because when you're creating a site you have to do the research and then decide what information you need and what information you can leave out because it's either not relevant or it's too complicated. But in the long run you do the elimination process with the information that you get, so it's really interesting and makes you learn stuff.

Other students commented on the authoring process of their collaborative knowledge construction activity. For example, some students were against the idea that students could make deletions on their wiki page. They felt it was preferable for their classmates to leave comments or suggestions, but not actually edit the original text. They felt this would give them the opportunity to refute changes made by others, and avoid what they described as "wasted work". According to one student: "I don't agree with people making deletions right away on your own thing. I think it's better that they leave a comment and say 'you could just do this' but you don't actually change the original text." Other students, however, felt comfortable about their classmates making changes to their wiki articles. In Jason's words:

> Well, I don't really mind [people editing my ecozone page] as long as I know that everything is completely recorded in the editing history, but I guess if it was just easier to undo the changes, like a bit more immediate, then it wouldn't bother people... but I guess when you're posting something to a wiki you have to know that someone else is going to make changes, so you just have to be okay with that and move on.

Overall, the students enjoyed making wiki artifacts as part of their biodiversity unit, and felt confident about using their peers' work as a resource for later activities. Since the wiki environment was new to students, however, future implementations of this activity would need to consider the social aspects of wiki authoring such as accepted norms for editing, work distribution and possible role assignment.

## Discussion and Conclusion

Although still ongoing, the findings from this research provide support for the Knowledge Community and Inquiry model as a mechanism for engaging students in collaborative inquiry. The blended use of wiki-based knowledge construction and scripted inquiry activities enabled students to make productive use of their co-constructed community resource. Through co-design, we were able to establish a successful research partnership with teachers, resulting in curriculum materials in which the teachers had a strong sense of ownership. We see this partnership as a positive step towards the "hybrid culture" described by Bereiter (2002), in which researchers and teachers rely on each other when working towards an educational objective.

This research also raises new questions concerning collaborative practices in a wiki environment. For example, despite the relative success of the knowledge construction activity, we still know little about students' individual practices when working in a wiki: Why do some students contribute more than others? Are all wiki transactions of the same educative value? And what are the optimal conditions for encouraging students to work with the content contributed by their peers? These questions open up an important dialogue for further investigation – and characterization – of collaborative knowledge construction and inquiry.

## References


Bereiter, C. (2002). *Education and mind in the knowledge age* (pp. 382-418). Mahwah, NJ: Lawrence Erlbaum Associates.

Bielaczyc, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with technology. *Journal of the Learning Sciences, 13*(3), 301-329.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141-178.

Brown, A. L., & Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education*, (pp. 289-325). Mahwah, NJ: Erlbaum.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology*. New York: Springer-Verlag.

Collins, A., & Halverson, R. (2009). Rethinking education in the age of technology: The digital revolution and schooling in America. New York, NY: Teachers College Press.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15-42.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179-201.

Drucker, P. F. (1959). *Landmarks of tomorrow.* New York: Harper & Brothers.

diSessa, A. A. (2000). *Changing minds: Computers, learning, and literacy.* Cambridge, MA: MIT Press.
Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences, 11*(1), 105-121.
Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*, 213-232.
Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107.
Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education, 41*, 397-420.
King, A. (1993). From sage on the stage to guide on the side. *College Teaching, 41*(1), 30-35.
Kuhn, D., Black, J., Keselman, A., Kapla, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction 18*(4), 495-523.
Linn, M. C., & Eylon, B. S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 511-544). Mahwah, NJ: Lawrence Erlbaum.
Linn, M. C., & Hsi, S. (2000). Computers, Teachers, Peers: Science Learning Partners. Mahwah, NJ: Lawrence Erlbaum Associates.
Lemke, C., Coughlin, E., & Reifsneider, D. (2009). Technology in schools: What the research says: A 2009 update [Electronic Version]. Cisco/*Metiri Group.* Retrieved February 26, 2010, from http://www.metiri.com/
McCandliss, B. D., Kalchman, M., Bryant, P. (2003). Design experiments and laboratory approaches to learning: Steps toward collaborative exchange. *Educational Researcher, 32*(1), 14-16.
Naidu, S., & Järvelä, S. (2007). Analyzing CMC content for what? Computers & Education, *Computers & Education 46,* 96–103
Roschelle, J., Penuel, W. R., & Shechtman, N. (2006). Co-design of innovations with teachers: Definition and dynamics. *Proceedings of the Biennial International Conference of the Learning Sciences* (pp. 606-612). Bloomington, IN.
Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pages 69-97). Heidelberg, Germany: Springer-Verlag.
Scardamalia, M., & Bereiter, C. (1996). Adaptation and understanding: A case for new cultures of schooling. In S. Vosniadou, E. de Corte, R. Glaser, & H. Mandl (Eds.), *International perspectives on the design of technology: Supported learning environments* (pp. 149-163). Mahwah, NJ: Lawrence Erlbaum.
Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Norwell, MA: Kluwer Academic Publishers.
Schwab, J. J. (1962). The teaching of science as enquiry. In J. J. Schwab & P. F. Brandweine (Eds.), *The teaching of science* (pp. 3-103). Cambridge, MA: Harvard University Press.
Slotta, J. D. (2004). The Web-based inquiry science environment (WISE): Scaffolding knowledge integration in the science classroom. In M. C. Linn, P. Bell, & E. Davis (Eds.), *Internet Environments for Science Education* (pp. 203-232). Mahwah, NJ: Lawrence Erlbaum Associates.
Slotta, J. D. (2007). Supporting collaborative inquiry: New architectures, new opportunities. In J. Gobert (Chair), Fostering peer collaboration with technology. Symposium conducted at the biennial *Computer Supported Collaborative Learning (CSCL) Conference,* New Brunswick, NJ.
Slotta, J. D. & Linn, M. C. (2009). WISE Science: Inquiry and the Internet in the Science Classroom. Teachers College Press.
Slotta, J. D., & Peters, V. L. (2008). A blended model for knowledge communities: Embedding scaffolded inquiry. *International Perspectives in the Learning Sciences: Cre8ing a learning world. Proceedings of the Eighth International Conference for the Learning Sciences – ICLS 2008* (pp. 343-350). International Society of the Learning Sciences, Inc.
Songer, N. (2006). BioKIDS: An animated conversation on the development of curricular activity structures for inquiry science. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 355-370). Cambridge, MA: Cambridge University Press.
Stahl, G. (2000). A model of collaborative knowledge-building. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences* (pp. 70-77). Mahwah, NJ: Lawrence Erlbaum.
Tapscott, D., & Williams, A. D. (2006). *Wikinomics: How mass collaboration changes everything.* London: Penguin Group.
White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1)3-118.

# Spatial Intelligence and the Research – Practice Challenge

Moshe Krakowski, Yeshiva University, moshe.krakowski@yu.edu
Kristin Ratliff, University of Chicago, krratliff@uchicago.edu
Louis M. Gomez, University of Pittsburgh, lgomez@pitt.edu
Susan Levine, University of Chicago, s-levine@uchicago.edu

**Abstract:** Spatial intelligence is deeply related to success in the STEM disciplines (science, technology, engineering, and math) yet teachers are unaware of the concept. This paper reports on a teacher, researcher, and designer collaboration focused on translating research shown to improve spatial thinking practice. This translation is supported through a spatial "toolkit" of activities designed to raise teachers' spatial intelligence awareness and improve spatial thinking practice.

## Introduction

In this paper we will describe the efforts of the Spatial Intelligence and Learning Center (SILC) to impact elementary school teachers' knowledge and incorporation of spatial content in the school day. These efforts face two interrelated sets of problems. The first is a central problem of education research: namely, how to translate research insights developed in the lab into classroom practices. The second relates to the unique nature of spatial intelligence: as a domain of research it has no named corollary in the school day, greatly exacerbating existing difficulties in bringing research to practice. Though spatial intelligence has important implications for success in the STEM disciplines (Science, Technology, Engineering, and Math), it is not itself an independent discipline. We have yet to find a school that has a class called "Spatial Knowledge"; rather, unexploited opportunities for spatial instruction are ubiquitous throughout the existing curriculum.

In the research literature models have been proposed to bridge the research/practice divide (e.g. Bryk & Gomez, 2008; Schoenfeld, 2009). These models describe a research effort that is vertically integrated and involves design collaborations between teachers and researchers. In this paper, we describe a design effort that substantially follows these approaches, but is modified to accommodate a domain in which the desired content is entirely unknown to teachers, and does not exist in the school curriculum. We do so by way of a spatial "toolkit" of activities, an innovation tailored to the unique circumstances of spatial intelligence research, by allowing teachers the flexibility to incorporate spatial elements in their curricula, without focusing on one specific domain.

## Spatial Intelligence in the Classroom

A central hypothesis of SILC is that particular experiences that occur early in life are important in laying the groundwork for spatial development. However, unlike other domains, notably language and literacy development, we currently know little about the preconditions for the development of robust spatial skills. SILC researchers have addressed this question through a number of studies examining the development of spatial language, engagement in spatial activities, and the relation of individual variations in these areas to the development of spatial skills themselves, along with a variety of domains of academic achievement.

Recent findings have shown that individuals who are more advanced spatial thinkers are more likely to major in STEM disciplines in college and to choose STEM career paths, even controlling for verbal and mathematical achievement levels (e.g. Hedges & Chung, 2008; Shea, Lubinski, & Benbow, 2001). Correlations specifically between visual-spatial ability and initial mathematical skills are even apparent as early as preschool (Robinson et al., 1996) and relate to middle and high school math achievement (Wolfgang, Stannard, & Jones, 2001). These relationships are not surprising, as spatial reasoning is central to many domains, not only explicitly in fields such as geometry, algebra, and calculus, but also as a foundational component in many domains (e.g., spatial reasoning is required for the construction of the nonverbal mental models that may underlie early arithmetical thinking and is critical to the ability to understand and create the physical models, diagrams, and visualizations used in fields such as science and engineering).

This reality highlights the importance of understanding the development of spatial thinking and its support. While it was long believed that spatial reasoning ability is biologically fixed, there is mounting evidence that training enhances spatial performance (Baenninger & Newcombe, 1989; Huttenlocher, Levine, & Vevea, 1998). Given this context, we would like to understand how the classroom environment and school content contributes to spatial reasoning, and perhaps more importantly, whether we can translate the rich spatial intelligence data we gather in the lab to classroom settings. Doing so would allow us to modify the classroom environment to enrich spatial input, which may enhance the spatial capacities of all students, and might perhaps provide particular benefit for those students who enter school with deficits in spatial reasoning.

## Initial Pilot Work: Identifying a Problem

That we need to modify the existing classroom context to enrich spatial input was made clear in an early pilot study. In this work we explored the spatial environment of a cohort of eight teachers' classrooms in an elite private elementary school. We began by cataloguing the representations found on the walls of the classrooms, as representations have the potential to contain a great deal of spatial information, such as the information contained in graphs, charts, maps, and measurement tools (e.g. rulers and thermometers, as in Figure 1). Because representations represent content visually and include explicit spatialization of data, they constitute an obvious opening for teachers to target spatial abilities, yet our pilot study demonstrates this is rarely the case.

Each classroom was photographed, and the representations were entered into a database where they were organized by representation type, subject area, and spatial content, among other factors.

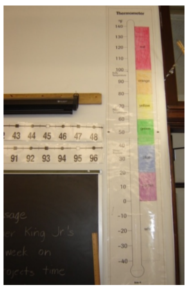


Figure 1. Thermometer Representation

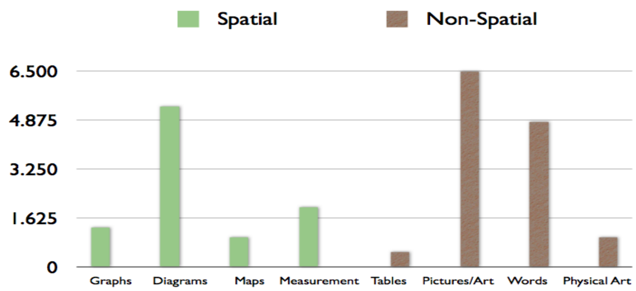



Figure 2. Average Number of Representations Per Classroom

Overall, there were more non-spatial representations than spatial representations (with an average of 12.9 per class to 9.7). As represented in Figure 2, diagrams constituted the only category of spatial representation frequently seen in classrooms. The majority of representations were decorative, displays of student work, or classroom guidelines, and very few required the use of spatial thinking (such as mental rotation, scaling, orientation, alignment, etc.), as even the diagrams (which do represent information spatially) did not make large spatial demands on students.

We followed up the representation catalogue by conducting clinical interviews with teachers in these classes. We tried to assess the degree to which they were aware of spatial intelligence and the degree to which they incorporated spatial elements in their classroom activities. For example, we gave teachers a sorting task, presenting them with more than 40 different pictures of real classroom representations. Teachers were asked to choose which of the assortment they would put on their classroom walls. We asked them to describe their decision-making process and to think about how they would utilize the representations. The following teacher's opening remarks as she sorted were typical of the responses we received:

> As I'm looking I'm sorting into what different programs require as resources for curriculum– that would be one pile. And included in those resources would be some things that implement our program… I would want to include things from different academic areas.
> I also think its really important to include children's work on display, some of these are hints about what to do when you're stuck… how to be prepared for school, I think it's important to have tips for kids, I also think in many of the classrooms there are class rules that have been generated by the class—those I like to post.

Over the course of the interview we pressed teachers to talk about representations that had explicit spatial content, such as the more spatially demanding charts and diagrams, and even in those cases teachers did not discuss spatial thinking. When asked to provide rationales for rejecting representations in their classes, teachers mostly cited aesthetics and the explicit relationship of the material to current curricular content (e.g. "too cutesy" or "low applicability to my class"). Thus, our pilot study demonstrated that even with respect to representations, which represent content visually, and often incorporate explicit spatialization of content, the teachers in our study rarely considered spatial elements or opportunities for spatial instruction within the existing curriculum

## Translating Lab Research

In the lab, researchers have had numerous insights into spatial learning, which, if translated to the classroom, could have an impact on students' spatial thinking. The problem, however, is that one cannot unproblematically translate lab research to the classroom. Even if we assume that our lab studies produce real insight that is relevant to classroom instruction, we must still struggle with the translation of that insight into something that has educational utility, a recognition that helped motivate the early development of design research (Brown, 1992). This is a problem that has been well articulated in the literature, as researchers have noted that laboratories are radically different environments than classrooms. In laboratory environments there is one-on-one interaction, with no outside noise or distraction, the intervention is heavily scripted so that experiments can be replicated, and the intervention is not embedded in any curricular context. Furthermore teachers are not researchers. Researchers develop deep expertise on very narrow problems, giving them the ability to devote time and energy to relatively small parts of the curriculum – a luxury teachers do not have. Often, teachers to not understand the interventions, or are unwilling to make changes to their practices, and in many cases they are deeply constrained by state, district, or school regulations.

A second set of problems is that, unlike other collaborative design projects, we needed to introduce teachers to entirely new concepts regarding intelligence, while having no pre-existing coherent curriculum to manipulate, improve, or change within the school curriculum. Improving and engaging students' spatial reasoning is important for a wide range of school content and skills, but it is not itself a learning domain; rather, it is an underutilized cognitive ability. Moreover, our pilot work indicated that teachers were entirely unaware of spatial intelligence as a concept. This challenge magnifies and makes more intractable some of the existing challenges to moving lab research into the classroom. The problem of acontextuality becomes much more profound when the lab studies do not concern themselves with school content, but with, for example, spatial alignment, and the problems of teacher impact become much more difficult, as well. Teachers do not have a baseline understanding of what spatial intelligence is, the way they do with content areas such as math, science, and reading. While it is true that many educational innovations are previously unknown to teachers, here we are not restructuring practices, reorganizing content, or devising new methods, but in essence raising awareness of an entirely new domain of intelligence, and asking teachers to act on that awareness.

To understand how these problems are magnified by the lack of content domain let us contrast them with substantial inquiry learning efforts. Early research, much of it based on lab studies, indicated that the deepest and most powerful learning takes place in context, and that acontextual learning was weaker and less meaningful. This insight (much simplified here) has led to a spate of programs to build curricula organized around meaningful activities, and more authentic organization of content. However while such programs may require pulling content from different areas of the traditional curriculum in order to organize that content along particular design principles, the basic content itself remains familiar. With spatial intelligence, however, there is no content that can be pulled from one place and reorganized; rather, many different content areas need to be taught with more and different information attached, and in some cases, in different ways.

## Overcoming the research – practice gap

In recent years, the literature has once again addressed the problem of the research-practice gap, and a growing body of literature has put forth a model of research to practice that addresses many of our first set of problems. This literature comes in response, however, to a different set of problems, posed by policy dictates that prioritize large randomized trials of educational interventions, and push for an understanding of "what works" in education.

In response to these mandates, a number of researchers (e.g. Bryk & Gomez, 2008; Burkhardt & Schoenfeld, 2003; Schoenfeld, 2009) have argued that in order for research to impact practice coherent institutions need to be developed that support multiple phases of development, from early trial research to broader testing, and finally to large controlled studies. They have pointed out that in other well-established fields a significant amount of research and development takes place well before any randomized trials begin, and that any randomized test of an intervention needs to be built on a solid theoretical base. In this literature design research is identified as an exemplar of the type of initial research that can help build theory, identify both *what* works and *why* it works, and can begin to flesh out the conditions under which interventions succeed and fail in practice. Accounts of this type share many similarities. They call for a vertical integration of the research effort, so that there is real coordination between the research and practice, a less fragmented infrastructure for research and development, with significant funds directed at the early, theory building stages, and a variety of research to practice models that begin with the iterative development of working prototypes and end with wide scale adoption and testing.

These models of a unified research effort that moves from existence proofs and design testing to widespread scaled-up interventions in schools also address the major problems with translating lab research to the classroom articulated above. Iterative, collaborative efforts between designers, teachers, and researchers are ideal for producing classroom interventions that work, as these collaborations, with focused effort, can overcome the gap between the lab and classroom, moving incrementally from one-on-one, scripted, a-contextual, researcher lead studies to dynamic classroom activities. This is necessarily an incremental process because the translation is so complex; it is rare that an intervention will be successfully translated in one session. Instead, multiple iterations take place, with teachers involved in the design and analysis, allowing the entire group to adjust to teacher difficulties, district mandates, and any other obstacles that may present themselves.

In our preliminary effort to move some of our spatial intelligence insights to the classroom, we envisioned a research to practice pipeline built on strong collaborations with researchers, designers, and teachers, much like the models described above. However, our second set of problems, that is, the unique constraints arising from the lack of target domain in the school day, and the total lack of teacher awareness of the concept of spatial thinking, forced us to modify the design effort to accommodate the special nature of the domain. In the next section we describe our design process, and how that design provided a flexible solution to these constraints.

## Resolving the curriculum gap: the spatial "toolkit"

To resolve these problems we adopted an approach that utilized the major collaborative features of an integrated research effort, while deliberately maintaining a broad flexibility that allowed us to work within existing school curricula and activities. Putnam and Borko (2000) describe two teacher support efforts, the SummerMath program and the CGI project, which combined intensive summer workshops with the support and involvement of researchers throughout the school year. Importantly for us, these projects were intended to change teachers' conceptions as well as their practices. Along those lines, we designed a summer work circle, bringing together nine lower-elementary public school teachers along with SILC researchers for intensive learning and design sessions, followed that up with sessions and support throughout the school year, and repeated the process again the next year, with 4 original teachers and 10 additional teachers. The nature of the support that we provided, however, was tailored to the nature of the material we were working with, and primarily took the form of a spatial "toolkit".

In an influential article on culture, Ann Swidler (1986) argued for two models of cultural impact, active in different historical circumstances. In one model culture provides a unified system that directs and orient behavior, and in another culture simply shapes a repertoire of tools, "habits, skills, and styles", which impact action indirectly, as people pick and choose from the toolkit in ways that may be fragmented and sometimes even contradictory. The notion of toolkit of resources has been adopted to talk about resources for organizational routines (Spillane, Gomez, & Mesler, 2009), as people select from available organizational resources to create strategies for action. This notion had great appeal for us. Traditional models of change, which implicitly assume a unified conception that is introduced to teachers, cannot function when there is no place in the school day or curriculum to direct action. Instead, we felt that we could work with the teachers to create an actual, literal, toolkit of practices and activities that draw on the insights we have developed in the lab, but are flexible enough to accompany teachers existing curricula. By working with the teachers on a toolkit we felt that we would be able to increase teacher awareness of spatial intelligence and its role in the classroom, and to help teachers design spatially rich activities to compliment their existing curricula.

Over the course of the first work circle in the summer of 2008 teachers and researchers co-produced the first iteration of a spatial toolkit. In this toolkit, teachers utilized the knowledge that they acquired in research presentations and applied it to their curricula, merging their expertise with that of the researchers. The teachers had free range to develop spatially rich projects on any aspect of the curriculum and classroom experience. For example, one pair of teachers developed a community mapping activity where the students built a map of their community, starting with the school at the center, and moving outward to the students' houses and other local landmarks, while another pair of teachers developed a technique for representing the daily schedule using spatial representations that were proportional to the length of time for each component of the schedule.

In order to understand how teachers were implementing toolkit items generally, during the year we maintained contact with the teachers, as they filled out a daily checklist of their use of spatial representations and activities, and reported back to us on their use of the initial items in the toolkit. We found that of the representations that teachers indicated that they used, 72 percent were used in an explicitly spatial way.

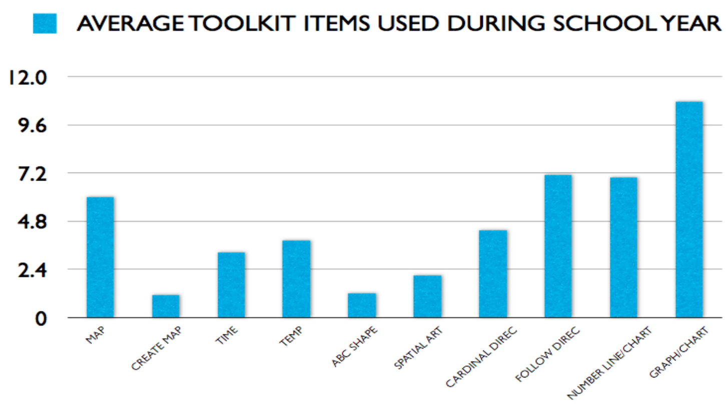



Figure 2: Distribution of Toolkit Items Used

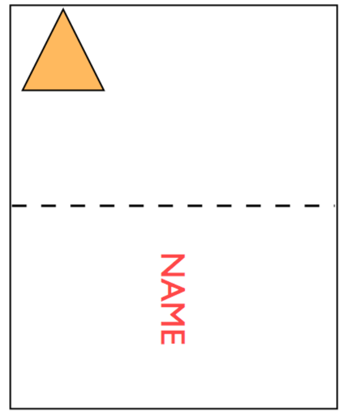



Figure 2a. Following Directions Example

We found that a domain did not need to be explicitly spatial for teachers to begin to incorporate explicitly spatial elements. One of the most frequently used toolkit items (Figures 2 and 2a, above) was an exercise that teachers use in the earliest grades to help students learn to follow directions. This activity could be done in any number of ways, but many teachers saw it as a perfect place for students to follow specifically spatial directions, such as folding a paper horizontally, putting a triangle in the top left corner, and writing their names perpendicular to the fold.

We also observed classes where the toolkit activities were being conducted. At points throughout the school year we held follow-up sessions where we had the opportunity to get additional feedback from teachers on the successes or failures of the toolkit activities. At these meetings we were also able to present teachers with new activity suggestions for them to tailor to their class use, based on our observations of their classes. For example, teachers expressed interest in a variety of puzzles, and SILC research has shown that puzzle play predicts children's performance on a mental transformation task (Cannon, Levine, Huttenlocher, & Ratliff, under review). We therefore developed a series of math puzzles (see Figure 3), such as a math domino game, a number grid puzzle ("What number am I?"), and a tangram math game, for the teachers to adopt in their classrooms.

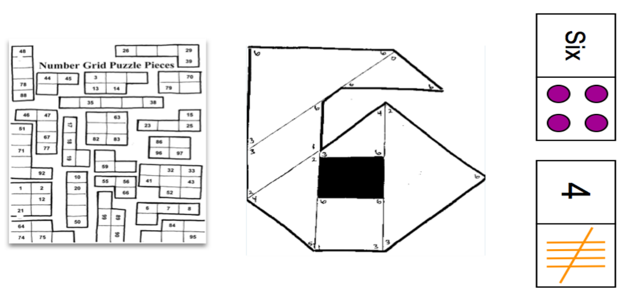



Figure 3. Math Puzzle Games

The toolkit format allowed us to modify and revise activities beyond the initial work circle, in order to accommodate teachers' needs throughout the year, and to insure that teachers were actively aware of spatial intelligence as a factor in their classrooms.

Teachers had the ability to pick and choose when and where the activities in the toolkit might be used, sometimes modifying them to fit their classroom needs. This design choice has numerous affordances particular to the material we were working with, but it also had some constraints. One constraint is that modified material sometimes changes such that the activity is no longer clearly representing the same concept as that in the lab activity. For example, a training study conducted in SILC labs (Levine, Kwon, Huttenlocher, Ratliff, & Dietz, 2009) demonstrated greatly improved ability to measure objects that are misaligned on the ruler when unit pieces are explicitly placed on the ruler (highlighting the space between hash marks as the unit, rather than the hash marks themselves), and when an item is first misaligned and then moved into alignment at the end of the ruler. In observing one 3rd grade class implementing a classroom version of this task, we noticed that the teacher had modified the activity to integrate with a number of different aspects of her curriculum. Instead of using the unit pieces to measure the spaces, and moving the measured item to the beginning of the ruler, she had the students use the units as part of a subtraction strategy, with some students putting units on all of the spaces up to the end of the item, and then removing those that did not align with the item. She explained this strategy to the students by saying "The turtle is four centimeters long because it ends at 6 and starts at 2; and 6- 2 is 4" (Figure 4).

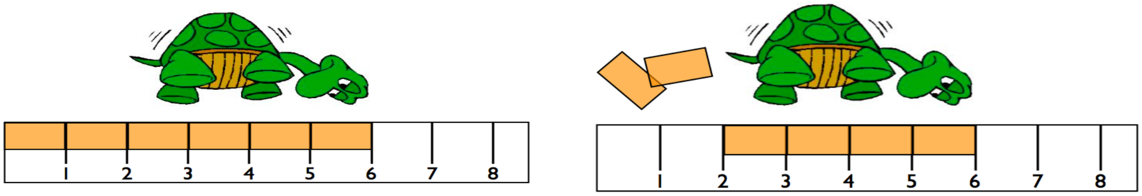
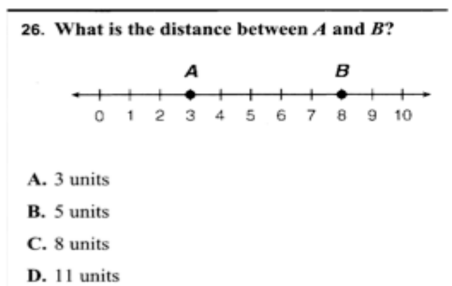


Figure 4: Classroom Adaptation of Measurement Training

This strategy seems like a good strategy for solving misaligned ruler problems, and certainly takes advantage of spatial processes such as alignment; however, it is not the strategy that we had investigated in the lab. On the other hand, when provided with a sample of the practice ISAT standardized assessment that the students take in 3rd grade, we noted that there were explicit number line questions that demanded exactly the subtraction strategy that the teacher was urging the students to utilize (Figure 5). This example highlights the sort of interplay between research and practice the toolkit approach allows.


**26. What is the distance between *A* and *B*?**



**A.** 3 units
**B.** 5 units
**C.** 8 units
**D.** 11 units

Figure 5: ISAT Item Using Subtraction Strategy

## Multiple types of Toolkit Impact

We believe our spatial toolkit impacts teachers precisely because it does not attempt a unified curriculum to replace an existing curriculum; it is instead a product that contains a multiplicity of resources, the sum of which can impact teachers' understandings and practices as they choose those that work for them. As we built toolkit resources together with teachers and observed their classes we were able to see five different ways in which spatial resources were integrated into classes, each of which represents some aspect of growing teacher awareness of, and interaction with, spatial intelligence in the classroom.

**Coherent Independent Content:** Although the toolkit as a whole does not offer a unified curriculum some activities do independently mirror content that exists in the school day and can be used as replacements. For example, using a ruler is a skill that is taught in the early grades, and, as noted above, is also reflected in the toolkit. Although some teachers modified the activity substantially, other teachers used the activity without very many changes. This is a hallmark of the toolkit approach, which allows teachers to choose how and when they will implement activities.

**Reorganized Content**: Much like an inquiry based curriculum will be organized around a driving question, pulling from numerous content areas in the process, some toolkit items take existing content and reorganize it around an underlying spatial concept. For example, together with teachers, we designed a series of activities that explored the use of the number line as a tool to understand the concept of unit as a basic foundation for understanding a wide range of math content, from addition and subtraction, to fractions and decimals, to area and perimeter. In the toolkit, teachers were able to utilize these activities as a coherent unit, if so they desired.

**Standalone Games and Attachments:** One of the great strengths of the toolkit is the presence of numerous items that are flexible and compatible with a wide range of classroom environments and opportunities for use. For example, teachers have used spatial puzzles and dominos during free time, center time, and as attachments to math lessons.

**Classroom Methods**: Much of the understanding that we hoped to develop in the teachers did not relate at all to particular content, but to techniques and tools that draw on spatial resources. For example, during the work circle teachers learned about research regarding the use of gesture (gesturing helps children understand spatial translation, (Ehrlich, Levine, & Goldin-Meadow, 2006)) and spatial analogies (Gentner, Levine, Dhillon, &

Poltermann, 2009). These can be used in any context in the classroom, and their use reflects a deep awareness of the utility of spatial reasoning a range of circumstances.

We observed a notable example of this in one of our second grade teachers' classes during the 2008-2009 school year. During the work circle we discussed research by Gentner that demonstrated that differences between two items are more easily discerned by students when the two items are highly aligned – that is, are highly similar, and differ only in one area. In a unit on time and clocks the teacher, unprompted by the researchers, took this concept of high alignment and applied it to the classroom to show students the correct orientation of the hands on the clock. Holding up two clocks that were supposed to represent 3:30 with nearly identical configurations, she asked them to identify what was different about the two clocks. On one clock the hands were at the six the three, and on the other clock the hands were at the six and between the three and the four (Figure 6). This was a prime example of a teacher using a high alignment spatial analogy to communicate content in the classroom.

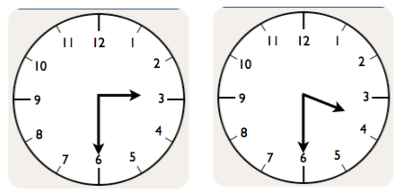



Figure 6: High Alignment of Clocks

**Language:** A major aspect of teacher uptake was in the use and incorporation of spatial language into the school day. This is one area that teachers immediately "got" and found ways to insert more, and more advanced, spatial language into their lessons. One teacher started using the cardinal directions when giving instructions to her students, and informed us during a feedback period that her students were no longer getting lost in the hallways of the school. Other teachers started to use more advanced terminology than they would have otherwise (for example: parallel, horizontal, diagonal, and symmetry were all mentioned by teachers).

## The Toolkit Approach as a Way to Change Teacher Thinking

In putting together a spatial toolkit we have been trying to introduce the concept of spatial intelligence to teachers, and to have teachers begin incorporating their understanding of spatial intelligence into their practices. Unlike a design for a new science curriculum or a new method for teaching fractions, many of the items in the toolkit might appear trivial—puzzles, games, following directions—but they serve an important purpose, and communicate a great deal of information. Because the activities are flexible, engaging, and teacher designed, teachers' perceptions of teaching have changed to include their awareness of spatial thinking as an important element of classroom instruction. Without any radical restructuring of the curriculum or of their methodology, teachers have been thinking more subtly about spatial content as they teach, and have been organizing their teaching around this thought process.

For example, teachers responded to an evaluation survey by uniformly commenting on how much their awareness had changed over the course of the study. One typical teacher wrote, "I have become much more aware of spatial intelligence and how to incorporate these things in my classroom". Nearly every teacher had some comment on the need to think explicitly about spatial intelligence while teaching (e.g. a sample of responses to "challenges I face": Being <u>very</u> explicit in spatial instruction; Thinking spatially about the lessons; Being aware of what I have in my classroom that ties into spatial thinking; Using specific spatial language, etc.). These responses reflect a reality in which teachers are using the toolkit as a resource that enables them to keep spatial intelligence in mind while constructing lessons and designing activities. This represents a major first step in impacting teacher practice.

## Final Thoughts

Spatial thinking is critically important for problem solving across the various sciences, as well as in many areas of daily life. Given recent evidence that spatial cognition is malleable, and hence that spatial learning can be fostered by effective technology and education (see Liu, Uttal, & Newcomb, 2008), we would naturally like to extend these findings to educational settings. The first steps in this process are making teachers aware of the existence of spatial intelligence, and impacting their practice to include activities that promote spatial reasoning.

In this study, we have tried to demonstrate some of the successes and challenges in introducing spatial intelligence to teachers and implementing change in the classroom. Because the concept is unknown to teachers, many areas that contain natural spatial learning opportunities are lost, and because spatial reasoning is not a subject taught in school we cannot easily build curricula to promote it. Many research-to-practice models encounter obstacles when there is no content in the school day matching the research in the lab. The development of the spatial toolkit helps fill in the gap, by increasing teacher awareness, and providing teachers with activities to supplement (rather than replace) their existing curricula. The toolkit approach allows teachers to pick and choose against a backdrop of resources and support, adjusting their understanding over time, and in concert with researchers.

As part of an ecology of events the toolkit can be an important early-stage product to spur the conversation between research and practice. After a number of work circle iterations we should have a good idea of what works and what doesn't, and under what circumstances different spatial activities will fit into the school day. A publishable end product could be produced and tried out more widely, giving a broader set of teachers the opportunity to build awareness of spatial reasoning through engagement in these activities. As teachers become more familiar with the concept, and more substantially adjust their teaching practices to include spatial content we can begin to test the impact of that content on students. Once we are certain that teachers "get it", in the sense of flexible integration with practice, we can explore how that translates to the students, and with that knowledge design even more targeted toolkit items to deepen student spatial ability.


Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*(5), 327-344.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141-178.

Bryk, A. S., & Gomez, L. M. (2008). Ruminations on Reinventing an R&D Capacity for Educational Improvement. In F. M. Hess (Ed.), *The Future of Educational Entrepreneurship: Possibilities for School Reform*. Cambridge: Harvard Education Press.

Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: Toward amore useful, more influential, and better-funded enterprise. *Educational Researcher, 32*(9), 3.

Cannon, J., Levine, S. C., Huttenlocher, J., & Ratliff, K. R. (under review). Early puzzle play: A predictor of preschoolers' mental rotation skill. *Developmental Psychology*.

Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental psychology, 42*(6), 1259-1268.

Gentner, D., Levine, S., Dhillon, S., & Poltermann, A. (2009). *Using structural alignment to facilitate learning of spatial concepts in an informal setting*.

Hedges, L. V., & Chung, V. (2008). *Does spatial ability predict STEM college major and employment: An examination of two longitudinal studies.* Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.

Huttenlocher, J., Levine, S., & Vevea, J. (1998). Environmental input and cognitive growth: A study using time-period comparisons. *Child Development, 69*(4), 1012-1029.

Levine, S. C., Kwon, M., Huttenlocher, J., Ratliff, K., & Dietz, K. (2009, August, 2009). *Children's understanding of ruler measurement and units of measure: A training study.* Paper presented at the 31st Annual Meeting of the Cognitive Science Society, Amsterdam, The Netherlands.

Liu, L., Uttal, D., & Newcomb, N. (2008). *A meta-analysis of training effects on spatial skills: What works, for whom, why and for how long?* Paper presented at the Conference on Research and Training in Spatial Intelligence, Evanston, IL.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4-15.

Robinson, N. M., Abbott, R. D., Berninger, V. W., Busse, J., Okamoto, Y., & Barbara, S. (1996). The structure of abilities in math-precocious young children: Gender similarities and differences. *Journal of Educational Psychology, 88*(2), 341-352.

Schoenfeld, A. H. (2009). Instructional research and the improvement of practice. In J. D. Bransford, D. J. Stipek, N. J. Vye, L. M. Gomez & D. Lam (Eds.), *The role of research in educational improvement* (pp. 161-188). Cambridge: Harvard Education Press.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93*(3), 604-614.

Spillane, J. P., Gomez, L. M., & Mesler, L. (2009). Notes on Reframing the Role of Organizations in Policy Implementation. *Handbook on Education Policy Research*, 409.

Swidler, A. (1986). Culture in action: Symbols and strategies. *American sociological review, 51*(2), 273-286.

Wolfgang, C. H., Stannard, L. L., & Jones, I. (2001). Block Play Performance among Preschoolers as a Predictor of Later School Achievement in Mathematics. *Journal of Research in Childhood Education, 15*(2), 173-180.

# Personal Beliefs about Learning and Teaching: Comparison of student teachers in the sciences and humanities at different stages of their studies

Natalia Schlichter, University of Goettingen, DFG Graduate Research Program 1195 "Understanding and Enhancing Educational Fit in Schools", Waldweg 26, 37073 Goettingen, Germany, Email: Natalia.Schlichter@sowi.uni-goettingen.de
Rainer Watermann, University of Goettingen, Department for Education, Waldweg 26, 37073 Goettingen, Germany, Email: rwaterm@uni-goettingen.de
Matthias Nueckles, University of Freiburg, Department of Educational Science, Rempartstrasse 11, 79098 Freiburg, Germany, Email: matthias.nueckles@ezw.uni-freiburg.de

**Abstract:** The present project deals with teachers' beliefs as implicit, steady assumptions about learning and teaching and especially with 'participation', as the least investigated. This exploratory study analyzes (1) the representation of beliefs regarding 'participation' in student teachers' views, the impact (2) of the stage of their studies, and (3) of their discipline studied on the beliefs. Student teachers were interviewed regarding their views on learning and teaching (*N* = 30). The content analysis shows different multifaceted beliefs. The results reveal an alteration in the beliefs during the studies: At the beginning there is a 'transmissive-emotional' focusing which changes progressively to a more cognitive focus including 'construction' and 'participation', and still 'transmission'. Furthermore, the internal structure of 'participation' in the views of the interviewees does not correspond with the theoretical concept. On the basis of these results the possibility of developing a questionnaire for closer investigation of 'participation' is being discussed.

## Introduction

Teachers' beliefs have become a focal point of research in the hope of gaining a better understanding of teacher behavior. Furthermore, since the rise of cognitive psychology, an more and more attention has been paid in particular to the juxtaposition of two opinions (Anderson, 1997; Greeno, 1997): the view of teaching as a 'direct transmission' (Resnick & Hall, 1998), with learning seen as an 'acquisition' of knowledge; and a 'cognitive construction' view of learning. The current research shows typical patterns: 'transmission' vs. 'construction' views (e.g., Staub & Stern, 2002). In addition we find 'acquisition' vs. 'participation' metaphors (Sfard, 1998), and 'content-centered' vs. 'learning-centered' teaching (Kember & Kwan, 2000). Empirical research so far concentrates on the 'transmission' vs. 'construction' views (Dubberke, Kunter, McElvany, Brunner, & Baumert, 2008); 'participation' is still being neglected. According to Sfard (1998), the role of 'participation' in the learning process, which was introduced as 'situated learning' (Greeno, 1997), should be considered to a greater extent in research.

Therefore, the current project centers on teachers' beliefs in general and on 'participation' in particular. This study is an exploration of the self-reflected views of learning and teaching as they emerged in interviews with student teachers. The purpose of this paper is to create a better understanding of student teachers' beliefs regarding participation. First, the theoretical framework will be presented, defining the term "teachers' beliefs" and showing the relevant research as well as different views of learning and teaching. The research questions of this study will then be presented. Next, the methodological information will be described. Finally, the findings will be presented and the possible implications will be outlined.

## Theoretical framework

### The term "teachers' beliefs"

Beliefs are subjective and experience-based perceptions of learning and teaching. Martinez, Sauleda, and Huber (2001) use the metaphor 'blueprint of thinking' to illustrate the type of influence exerted by beliefs on perception and behavior. Beliefs about teaching are well established by the time a student gets to college (Pajares, 1992); they do not change easily or rapidly and are not independent of the discipline taught (Boulton-Lewis, Smith, McCrindle, Burnett, & Campbell, 2001; Staub & Stern, 2002; van Driel, Bulte, & Verloop, 2007).

'Belief' is still a 'messy construct', as Pajares (1992) called it. Among the terms used to describe teachers' thinking we find: beliefs, conceptions, orientations (Kember, 1997), perspectives, intentions, teaching approaches (Trigwell, Prosser & Taylor, 1994), and teaching perspectives (Pratt, 2002). Because of largely congruent contents all these work will be equally employed for the present study.

We define teachers' beliefs as implicit, steady, experience-based assumptions about the nature of learning and teaching: about the learner, the teacher, the learning and teaching process, instruction, classroom management, the genesis of knowledge and its growth.

## Research on teachers' beliefs

Before addressing the relevant studies, the theoretical background of different beliefs should be briefly outlined. As pointed out by Sfard (1998), the theoretical paradigms of learning can provide a suitable frame of reference on teachers' beliefs. The author differentiates between two basic metaphors: 1) individual acquisition of knowledge and 2) participation in the community of practice. Here, learning is regarded as a process of becoming a member of a certain community, which requires the ability to communicate in the language of the community and to act according to its particular norms. The learner is interested in participating in various activities rather than accumulating personal possessions. In contrast to this position, the individual acquisition metaphor considers learning as the result of a transmission process of information from the teacher to the student. Sfard (1998) describes both metaphors under the following aspects: goal of learning, learning, student, teacher, knowledge, knowing, and motivation. Kember (1997) also employs several dimensions to delimit the conceptions of teaching: teacher, teaching, student, learning, content, knowledge.

Sfard (1998) assignes behaviouristic and constructivistic notions of learning to the first metaphor and notions of situated learning to the second metaphor (Bruner, 1996; Greeno, 1997; Lave, 1987; Lave & Wenger, 1991). Contrary to this position, Martinez et al. (2001) differentiates three main metaphors: 1) learning as transmission of knowledge, 2) the constructivist metaphor, and 3) teaching and learning as a social process. According to the constructivist metaphor, learning is described as an active, constructive process which is influenced by the subject's prior knowledge and takes place in particular contexts (e.g., Bereiter, 1994; Staub & Stern, 2002). According to Trigwell and Prosser there is also the distinction between 'conceptual-change/student-focused' and 'information-transmission/teacher-focused' approaches of teaching (see also Norton, Richardson, Hartley, Newstead, & Mayes, 2005; Postareff & Lindblom-Ylänne, 2008; Trigwell, Prosser & Taylor, 1994).

Of equal interest is the research that takes into consideration more than two opposing beliefs: for instance Pratt, Collins, and Jarvis-Selinger (2001) and Alger (2009). Pratt et al. (2001) worked out five teaching perspectives: 1) transmission: Effective teaching requires a substantial commitment to the content or subject matter; 2) apprenticeship: Effective teaching is a process of students enculturation into a set of social norms and ways of working; 3) developmental: Effective teaching must be planned and conducted "from the learner's point of view"; 4) nurturing: Effective teaching assumes that long-term, hard, persistent effort to achieve comes from the heart, as well as the head; and 5) social reform: Effective teaching seeks to change society in substantive ways.

There is a growing consensus in research that teachers' beliefs affect the learning of the students. Teachers' beliefs influence their own perceptions, which in turn affect their practices in the classroom (Calderhead, 1996).

Empirical studies have already shown that the 'cognitive construction' view, which deals with the cognitive activation and constructive support of the learners, seems to be more promising in promoting a deeper understanding during students' learning than the 'direct transmission' view (Aquirre & Speer, 2000; Dubberke et al., 2008; Staub & Stern, 2002). In contrast to this, the 'participation' view has been addressed less, and there are no suitable questionnaires to measure teachers' beliefs about learning and teaching in terms of 'participation'. Therefore, in this study, we focus on the 'participation' aspects of teachers' beliefs.

## Research questions and hypotheses

We designed the present study to answer three main questions:

1. To what extent are the beliefs about learning and teaching described above represented in the views of student teachers?

It is necessary to find out more about what the student teachers' views of learning and teaching are in general and how important they regard 'participation' in the learning process in particular.

2. Are there any similarities among the student teachers at the same stage of their studies?

We want to compare the beliefs of the student teachers at the beginning, in the middle and at the end of their studies. This question should provide more insight into the assumption that the increased knowledge and practical experience which the students gather during their studies results in more sophisticated beliefs about learning and teaching.

3. Are there any similarities among the student teachers in the related disciplines?

This question should illuminate whether there are any similarities among the specific discipline communities. It can be assumed that differences will exist between the beliefs of student teachers in the sciences and student teachers in the humanities.

Consequently, the current study was designed to investigate the beliefs of student teachers with the aim of highlighting their opinions regarding the role of 'participation' in the learning process. Moreover, if the answers to the latter two questions should yield converging results, the design would provide an insight into the structure of beliefs of student teachers dependent on their knowledge and practical experience, related to the specific discipline.

# Method

## Sample and Design

The sample consisted of a total of $N = 30$ interviewees (11 male, 19 female, mean age: 23.7 years), who were all students undergoing their teacher preparation program at the University of Goettingen. Two discipline communities were represented in the sample: a) the sciences, e.g. mathematics, physics, chemistry, biology and geography ($n = 15$); and b) the humanities, e.g. languages, history and theology ($n = 15$). Furthermore, the participants were at different stages of their studies: a) at the beginning ($n = 10$, mean age: 20.9 years), b) in the middle ($n = 10$, mean age: 23.1 years), and c) at the end ($n = 10$, mean age: 27.2 years).

## Inquiry Method

For the current study, we used the methodology described by White and Gunstone (1992). We elicited the participants' views on learning and teaching by asking them to draw a picture which showed their concept or theory of learning and teaching. The participants were invited to the interview without much prior information about the topic of our study. The intention was to let them communicate their own thoughts with as little external influence as possible. After about 15 minutes of drawing, we asked the participants to explain the picture, thus describing their concepts of learning and teaching. When they had finished doing so, we asked them pre-designated questions on certain aspects that were of interest within our theoretical framework, like: Is the teacher or the learner at the center of the learning process? Who is more important in the learning process: the learner or the teacher? What is the role of the learner in the learning process?

The interviews were audiotaped and transcribed. In addition, in order to be able to analyze nonverbal information, we videotaped the participants together with their picture during the interview. All of the interviews were codenamed to ensure that the data (both verbal and videotapes) remained anonymous and inaccessible to third parties.

## Data Analysis

The method we used to analyze the data was based on Chi's "Quantifying Qualitative Analysis of Verbal Data" (1997). The author recommends the application of this method for verbal data such as explanations or reports. She emphasizes the main goal of this analysis method as understanding the representation of the knowledge. Further, we used the computer program MAXQDA, which supports the coding procedure of verbal data. As preparation for the coding, the interviews were first segmented into single coding units. For this purpose, we used a procedure originally suggested by Chi (1997). Each interview transcript was divided into segments so that each segment could be coded independently. The preferred segmentation technique was based on semantic futures, according to which the content of the utterances had to be interpreted to determine the segment boundaries, such as knowing that the topic of discussion has been changed. In general, it is a proposition or an idea. Mostly, one segment is one idea unit:

Example:

…and that I refer to my experiences that I gained when I was a pupil//

Furthermore, the granularity is referred to as a main point of analysis (see Chi, 1997). Grain sizes can vary from coarse to fine. For our interview analyses, we opted for the fine grain. Consequently, one segment is limited to one sentence, as in the above examples. However, the structure of the speech was often unclear and disorganized; sometimes, it took more than one sentence to express an idea to its conclusion. In this case, one segment might have been longer than in other interviews:

Example:

…for this purpose, I thought to myself, it should be some sort of old-fashioned seating arrangement, for example one pupil sitting on a bench at a table, then the next row, then the next, and the teacher standing here at the front holding the pointer in his hand and teaching at the blackboard//

The first step was to develop a deductive coding scheme. Therefore, we used aspects as 'goal of learning' or 'knowledge' (e.g., Kember, 1997; Sfard, 1998). The coding process was performed by two raters who were trained for this coding procedure. After discussing the category system, they coded the interviews independently. The inter-rater reliability as determined by Cohen's Kappa was good ($\kappa = 0.82$).

The results of this coding process revealed a quite accurate correspondence with the categories of the "Teaching Perspectives Inventory" (TPI) according to Pratt et al. (2001). In the following section, we present these categories as well as the results of the study.

# Results

## Category System

Table 1 shows the category system. Due to space restrictions, the description of each category cannot be presented in complete detail here. The table presents a small excerpt of all of the aspects from each category.

Table 1: The category system.

| Name of the category | Content description of the category |
|---|---|
| Transmission | • learning as gaining achievement and competencies by the individual<br>• teacher as an instructor of learning<br>• knowledge as substance and property by the individual |
| Nurturing | • productive without fear of failure<br>• teacher's responsibility to create an atmosphere without fear<br>• the importance of self-efficacy or self-esteem of the pupils |
| Participation | • learning as a process of enculturation of students into a set of social norms and ways of working<br>• teacher is an expert in his/her discipline and a moderator and supporter of learning processes |
| Construction | • learning as an active and constructive process, depended on the previous knowledge<br>• teaching from the learner's point of view<br>• suitable methods: experience-based and discovery-based learning |
| Social Reform | • collective as an object of teaching rather than individual<br>• encouragement of learners to take a critical stance<br>• learning as developing a world view and opinions |

To demonstrate two of these categories, the subcategories with anchored examples will be presented in the following:

> Subcategory of 'nurturing' – well-being:
> The school and the lesson must take place in a protected space, where the teacher creates emotional safety so that the children feel happy and comfortable and do not have any fears regarding learning.
> Subcategory of 'participation'- 'supporting class community':
> As a pillar of the classroom community, the teacher supports the development of the perception of the class as one community. He/she has to ensure that the pupils act as a group and not as individuals.

## Supplementary Categories

Also of interest are the categories that do not concern the theoretical learning paradigms, but rather address some supplementary influences on interviewees' thoughts during the interviews. For instance, there were ideas about learning, which could be described as naïve positive attitudes, for example about the effects of some specific teaching methods, as well as idealistic attitudes about the teaching profession.

> …so we can make the lessons very, very interesting if we use the new and creative teaching methods//
> The smaller the group, the better the learning//

The interviewees addressed their experiences as learners very often, but described their experience as teachers surprisingly rarely, although most of the student teachers were undergoing their school-based teaching practice during their studies. The focus of the present study is not on such aspects, so we will not address them further in our analysis.

## Internal Structure of Participation

In the theoretical presentation of the construct 'participation' (e.g., Sfard, 1998), there are many aspects that clarify the understanding of learning and teaching according to 'participation'.

The interviewees in the present study made their statements spontaneously. Consequently, it is conceivable that some aspects of their beliefs were not addressed. Hence, the view of the interviewees regarding 'participation' differs from the theoretical construct. For instance, some of the theoretical aspects were not represented completely. The figure below demonstrates the number of the interviewees who made some statements regarding the respective individual aspect (see Figure 1).

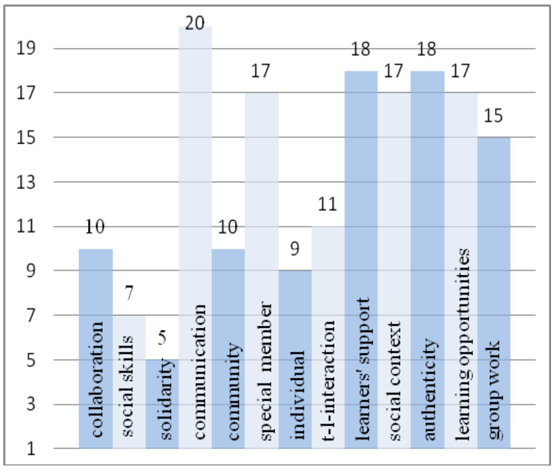



Figure 1.The internal structure of the 'participation'.

Most of the interviewees (*n* =20) commented on the communication process in the class. Communication, authenticity in learning, support by the learners, social context, the teacher as a special member of the class community, the teacher's responsibility to provide learning opportunities were also called often. Some other aspects as group work, interaction process close this list.

Of interest is also the fact that some aspects could be seen as overlapping with the category 'nurturing' because of their emotional value, for instance 'solidarity'. In conclusion, the first research question can be answered as follows: the beliefs of the student teachers in the present study contained five main beliefs and also some supplementary contents. Furthermore, the beliefs of interviewees contained 'participation', but the structure did not correspond to the theoretical construct. In the following section, a closer look will be taken at the different groups of the student teachers.

## Student Teachers' Beliefs

In total, the interviews yielded 2262 segments. 74.2 % of them were assigned to the main categories of beliefs about learning and teaching; the other 25.8 % were described in the section above (see Supplementary Categories). The following table (see Table 2) shows the means (in percent) and standard deviations for the values in the main categories for all the groups of student teachers.

Table 2: Descriptive statistics: Means (in percent) and standard deviations.

| | Whole sample (*N* =30) | Student teachers at the beginning of their studies(*n* =10) | Student teachers in the middle of their studies (*n* = 10) | Student teachers at the end of their studies (*n* = 10) | Student teachers in the sciences (*n* = 15) | Student teachers in the humanities (*n* = 15) |
|---|---|---|---|---|---|---|
| Total statements | 74.20 (29.13) | 73.30 (28.83) | 70.10 (20.76) | 79.20 (37.74) | 64.40 (30.21) | 84.00 (25.27) |
| Transmission | 25.77 (12.80) | 29.22 (17.34) | 23.93 (8.98) | 24.17 (11.26) | 26.07 (15.47) | 25.48 (9.99) |
| Nurturing | 15.18 (12.12) | 23.86 (12.85) | 11.01 (9.71) | 10.69 (9.40) | 13.47 (10.92) | 16.90 (13.37) |
| Participation | 23.9 (10.83) | 19.53 (7.39) | 29.85 (12.34) | 22.33 (10.37) | 25.70 (10.92) | 22.11 (10.81) |
| Construction | 19.25 (9.66) | 13.21 (9.68) | 21.35 (9.03) | 23.21 (07.91) | 27.30 (13.54) | 26.83 (11.94) |
| Social Reform | 03.65 (3.21) | 04.50 (3.34) | 01.10 | 0 | 00.12 (0.47)0.12 | 00.85 (2.24) |

## The impact of the stage of studies on the beliefs

The second research question concerned the different stages of the studies. Our sample contained students (a) at the beginning, (b) in the middle and (c) at the end of their studies to compare the influence of experiences from the studies on the beliefs. In order to examine the difference in each belief for each group, we conducted a multivariate GLM model (within subject design). The main effect of beliefs is significant: $F = 26.23$, $p < .001$, and reveals quite strong power, at $Eta^2 = .49$ as well as the interaction between beliefs and the stage of the studies: $F = 2.58$, $p< .05$, $Eta^2 = .16$. The findings are demonstrated in the figure below (see Figure 2).

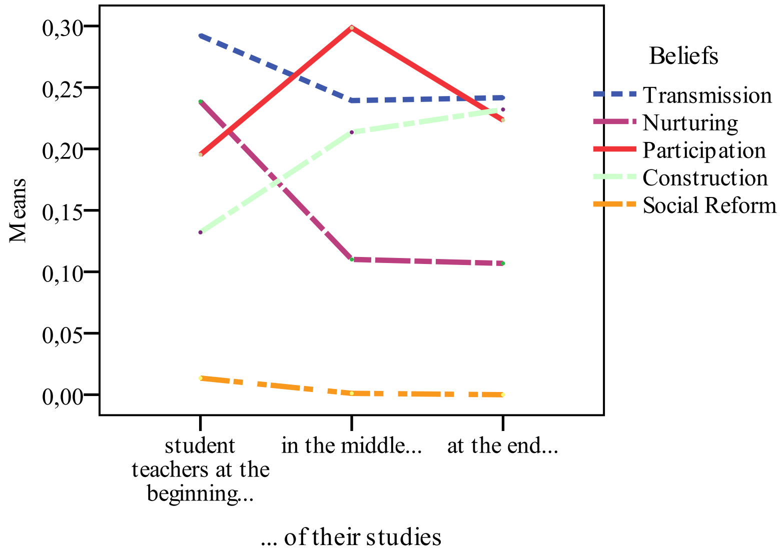



Figure 2. The distribution of the beliefs for each group of student teachers.

The difference between the student teachers at different stages in each belief was tested by means of univariate analysis of variance (ANOVA). The findings show a significant effect of the stages of the studies on the belief 'nurturing' at the .05 level: $F = 4.87$, $p < .05$, $Eta^2 = .27$. The test of contrasts shows a significant difference between the student teachers at the beginning and in the middle of their studies: $F = 23.25$, $p < .001$, $Eta^2 = .46$.

The difference in 'participation' is also significant. The effect of the stage of the studies on 'participation' is, however, significant only at the .1 level: $F = 2.71$, $p < .1$, $Eta^2 = .17$. The difference between the different groups of student appearances as follows: the student teachers in the middle of their studies show the biggest values compared with much lower values of the student teachers at the beginning and lower values of the student teachers at the end of their studies: $F = 62.11$, $p < .001$. This effect power is quite strong, at $Eta^2 = .70$.

Furthermore, the statistics reveal the differences in 'construction'. The effect of the stages of the studies on 'construction' is significant at the .05 level: $F= 3.57$, $p < .05$, $Eta^2 = .21$. The test of contrasts shows that the student teachers at the beginning of their studies have the lowest values in 'construction' compared with quite higher values of the student teachers at the end and also higher values by the student teachers in the middle of their studies.

For all the other beliefs, the differences were not significant. It is noticeable that in some categories, not all groups made statements: For 'social reform', none of the students at the end of their studies made any statements.

The following result could be already seen in the figure 2: The student teachers at the end of the studies possess the three beliefs 'transmission', 'construction', and 'participation' at the quite similar level; the belief 'nurturing' is also represented, but in noticeable smaller extend. This difference was tested by means of t-Test, computing for the beliefs 'transmission', 'construction', and 'participation' a common variable named 'cognitive focus' previously. The means (standard abbreviations) for 'cognitive focus' and 'nurturing' are respectively 23.23 (4.17) and 10.69 (9.40). This difference is significant at the level .05: $t = 3$ ($df = 9$), $p < .05$.

Due to the quite small groups of the students, all statistics were also reconsidered by means of non-parametric tests which carried out the findings like the above. Nevertheless, the conditions for ANOVA are fulfilled: the normal distribution and the homogeneity of variances are given.

To summarize, the student teachers at different stages of their studies have significant different beliefs. The development of the beliefs during the studies changes from the 'transmissive-emotional' focus including 'transmission' and 'nurturing' toward a 'cognitive' focus including 'construction', 'participation', and still 'transmission'.

### The impact of discipline on beliefs

To examine hypothesis 3, namely that the beliefs of student teachers in the sciences differ from those of student teachers in the humanities, we conducted analysis of variance (ANOVA), which proved to be insignificant. To sum up these final results, it can be stated that the expected differences between the beliefs of the student teachers from the two disciplines were not found.

## Conclusions and implications

In this paper, we described a study conducted recently at the University of Goettingen. It was designed with the aim to highlight how the different beliefs are represented in the thinking of the student teachers and what differences there are in these beliefs at various stages of their studies and in different disciplines.

At the beginning of the present study we did not have any expectations as to which beliefs would revealed by the interviews. In a first analysis, we expected two main paradigms such as 'individualistic' vs. 'social' paradigms in learning sciences (Sfard, 1998) or perhaps a 'transmission' vs. 'construction' view on learning and teaching (Staub & Stern, 2002). In the end we found not just one or two, but five different beliefs. These results could be seen in a positive light. An increasing number of researchers are calling for plurality in looking at the beliefs about teaching and learning. Pratt (2002) warns against the "one size fits all" notion of good teaching. Even Sfard (1998) points out the "dangers of choosing just one" of the two metaphors.

In keeping with the recent theoretical framework and empirical research that researchers and educators should pay more attention to the concept of 'participation' in the field of the learning science (Alger, 2009; Lave &Wenger, 1991; Sfard, 1998), our study set out to investigate student teachers' beliefs regarding 'participation' in more detail. In so doing, we analyzed the internal structure of the notion of 'participation' found in our interviews. This analysis showed that the understanding of learning as 'participation' by the student teachers in our study does not correspond to the theoretical concept. This can possibly be explained by the fact that the interviewees spoke spontaneously. Consequently, it is conceivable that some aspects of their beliefs were not addressed. The overlapping areas relate to emotional aspects of group processes like solidarity, which reveals that the future teachers value the idea of 'participation', not only from the perspective of advantages for learning but also in terms of the emotional advantages. In addition, many aspects concern social aspects of group processes. There are no didactic aspects or statements about specific learning processes in a 'learning community', with the exception of 'providing learning opportunities by the teacher'. The constitutive point of 'participation' – learning as enculturation – is lacking. Thus, this indicates that a one-sided view should be changed, and that the teacher education should be able to initiate such a change.

Furthermore, the present study confirmed the hypothesis of underlying similarities among the beliefs of the students at the same stages of their studies. But, the experimental design of the present study was cross-sectional, which is why the differences between the student teachers of different stages of their studies cannot be interpreted as definite changes in beliefs, but should at present rather be understood as hint thereof. Additionally, the teacher preparation program included not the same contents, because of the present system of Higher Education in Germany. Indeed, especially the clear preference of the student teachers in the middle of the studies for 'participation' emphasizes this fact. This especially preference may be a cohorts effect. In addition, the student teachers at the beginning of their studies understood learning primarily in terms of transmission of knowledge from the teacher to the pupil, seeing the main responsibility of the teacher in preparing and presenting of the contents of his/her discipline as well as creating an atmosphere without fear in order to secure of self-efficacy and self-esteem of the children. While the student teachers at the beginning of their studies show this 'transmissive-emotional' focus on the learning and teaching, the beliefs by the student teachers at the end of their studies tend toward the 'cognitive' focus. Thus, this alteration in the beliefs might be attributable to the increased knowledge the students gather during the studies and practical experience from the internships allowing to evaluate the role of teacher training a little deeper. We hope that these findings will be fruitful for educators who work in the field of teacher preparation program.

The sample also comprised student teachers from various disciplines. The recent research confirmed similarities in the beliefs in the same or relative disciplines (Boulton-Lewis et al., 2001; Norton, 2005; Pajares, 1992; Pratt et al., 2001; van Driel et al., 2007). Contrary to this research, in the present study differences between the student teachers in the sciences and the humanities were not found. In the Higher Education program in Germany there are two disciplines which must be studied. Thus, there are students in the sample who study for instance physic and language. The assignment of the students to each group was made according to the favorite discipline named by the student. This method of assignation seems to be not selective enough.

The aim of our study was to highlight the beliefs and how they are represented in the thinking of the interviewees. Restrictions were kept as more as possible. Due to the restrictive nature of questionnaires, we decided to work according to the inquiry method of White and Gunstone (1992). The next step could be to use the results of the present study to develop a questionnaire to measure 'participation' even more closely. Furthermore, a future study could interview other groups of people. Beliefs of experienced teachers seem to be of particular interest in the light of the present project. As a conclusion of the present study we must bear in mind that our investigation reveals not actual beliefs, but rather, in the best case scenario, the 'blueprints of beliefs' in the particular interview situation.

## References


Aguirre, J., & Speer, N. M. (2000). Examining the relationship between beliefs and goals in teacher practice. *Journal of Mathematical Behavior*, *18*(3), 327-356.

Alger, C. L. (2009). Secondary teachers' conceptual metaphors of teaching and learning: Changes over the career span. *Teaching and Teacher Education*, *25*, 743-751.

Anderson, J., Reder, L. M., & Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational Researcher*, *26*(1), 18-21.

Bereiter, C. (1994). Constructivism, socioculturalism, and Popper´s World 3. *Educational Researcher*, *23*(7), 13-20.

Boulton-Lewis, G. M., Smith, D. J. H., McCrindle, A. R., Burnett, P. C., & Campbell, K. J. (2001). Secondary teachers' conceptions of teaching and learning. *Learning and Instruction*, *11*, 35-51.

Bruner, J. (1996). *The Culture of Education.* Cambridge, MA: Harvard University Press.

Calderhead, J. (1996). Beliefs and knowledge. In *David C. Berliner & Robert C. Calfee (Eds.), Handbook of educational psychology* (S. 709-725). New York: Macmillan.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, *6*(3), 271-315.

Dubberke, T., Kunter, M., Mc Elvany, N., Brunner, M., & Baumert, J. (2008). Lerntheoretische Überzeugungen von Mathematiklehrkräften: Einflüsse auf die Unterrichtsgestaltung und den Lernerfolg von Schülerinnen und Schülern. *Zeitschrift für Pädagogische Psychologie*, *22*, 193-206.

Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher*, *26*(1), 5-17.

Kember, D. (1997). A reconceptualization of the research into university academics' conceptions of teaching. *Learning and Instruction*, *7*(3), 255-275.

Kember, D., & Kwan, K. (2000). Lecturers' approaches to teaching and their relationship to conceptions of good teaching. *Instructional Science*, *28*, 469-490.

Lave, J. (1987). *Cognition in practice.* Cambridge, UK: Cambridge University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Martinez, M., Sauleda, N., & Huber, G. (2001). Metaphors as blueprints of thinking about teaching and learning. *Teaching and Teacher Education*, *17*, 965-977.

MAXQDA, Software für qualitative Datenanalyse, 1989 – 2007, VERBI Software. Consult. Sozialforschung GmbH, Berlin-Marburg-Amöneburg, Deutschland.

Norton, L., Richardson, J. T. E., Hartley, J., Newstead, S., & Mayes, J. (2005). Teachers' beliefs and intentions concerning teaching in higher education. *Higher Education*, *50*, 537-571.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*(3), 307-332.

Pratt, D. D. (2002). Good teaching: One size fits all? *New Directions for Adult and Continuing Education*, *93*, 5-15.

Pratt, D. D., Collins, J. B., & Jarvis-Selinger, S. (2001). Development and use of the Teaching Perspectives Inventory (TPI). Paper presented at the Annual Meeting of the American Educational Research Association (AERA). Available on: http://www.teachingperspectives.com/PDF/development1.pdf (25.02.2010).

Resnick, L. B., & Hall, M. W. (1998). Learning organizations for sustainable education reform. *Daedalus*, *127*, 89-118.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, *27*(2), 4-13.

Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, (94),* 344-355.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grows (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: Macmillan.

Trigwell, K., Prosser, M., & Taylor, P. (1994). Qualitative differences in approaches to teaching first year university science. *Higher Education*, *27*, 75-84.

Van Driel, J. H., Bulte, A. M. W., & Verloop, N. (2007). The relations between teachers' general beliefs about teaching and learning and their domain specific curricular beliefs. *Learning and Instruction*, *17*, 156-171.

White, R. T., & Gunstone, R. F. (1992). *Probing understanding*. London: Flamer Press.

# A Longitudinal Approach to Appropriation of Science Ideas: A Study of Students' Trajectories in Thermodynamics

**Abstract:** This article provides an empirical analysis of data collected during the implementation of a teaching proposal on thermodynamics in a class of 20 students (17 year-olds) of a scientifically-oriented secondary school in Italy. During the activities, each student made evident progress in gaining intellectual autonomy as they took part in the teaching/learning dynamics of the classroom. Although the research had, up to now, an empirical orientation, this paper aims to provide a contribution for advancing theory development in physics education research. The study gives an example of the application of a specific model of teaching/learning (the Model of Longitudinal Development) that acted as "framework for action" in the design of the learning environment as "properly complex territory." The data analysis gives indications for how to develop a theoretical understanding of the concept of "personal learning trajectory" and it provides a basis for exploring the factors that can trigger intellectual autonomy.

## 1. Introduction

This article focuses on an empirical analysis of data collected during the implementation of a teaching proposal on thermodynamics in a regular 12th grade class in a scientifically-oriented secondary school in Italy.The study is an example of application of a specific model of teaching/learning (the Model of Longitudinal Development, described in Section 2), that acted as "framework for action" (diSessa & Cobb, 2004) in the design of a learning environment as "properly complex territory."

A framework for action is intrinsically and self-consciously heuristic. The purpose of a framework for action is to provide some useful focus and direction for the design of learning environments. However, because frameworks for action may involve multiple, diverse design elements and plans (some of which may have inexplicit underlying assumptions), frameworks for action are not themselves "theories of teaching and learning" that can be tested or falsified. The results of a design that is informed by the heuristic principles of a particular framework for action are iterative in nature, and often involve making explicit and gaining a better understanding of the initial choices made by intuition or guess-work and guided by previous design experiences. This is what happened also in this study: through feedback from our students, we progressively gained a refined awareness of both basic features and the potential of the materials we had designed. In sections 3 and 4, we document this progress by stressing what we had explicitly planned before and during the implementation and what became, to our surprise, evident in the monitoring of the classroom activities.

Although empirical in orientation, this study intents to go beyond purely heuristic findings and aims to provide a contribution to the enterprise of advancing theory development in physics education. The choice of focusing the study on the specific concept of "learning trajectory," known in the research literature of design studies (Confrey, 2006), did indeed allow us to find important hints in the data for developing the research from a more theoretical point of view. In particular, by working on real examples of learning trajectories constructed by the students themselves, the fairly abstract and mainly evocative word "trajectory" seems to gain specific and operatively recognizable properties. These properties appear to be good candidates not only for defining it, but also for potentially including it in a "local theory" (diSessa & Cobb, 2004), aimed at explaining the connections between individual and collective growth and individuating the factors that can trigger intellectual autonomy through the construction of personal learning trajectories.

Section 5 is devoted to implications for future research and stresses the need to point out such theoretical foci of attention on which future studies can be carried out.

## 2. The Model of Longitudinal Development (as a framework for action)

The Model of Longitudinal Development (MLD) is the main product of the Italian PRIN project *"F21 - Teaching and learning pathways in physics for the 21th century, in both "formal" and "informal" education"* (2004-2007). This project involved nine Italian research groups in physics education and was coordinated at national level by Paolo Guidoni.

MLD is comprised of a set of shared criteria for designing teaching materials that allow the cognitive potential of the pupils to be progressively exploited and tuned to the construction of physics knowledge along the pre-university curriculum (from Kindergarten to upper secondary school) (Guidoni & Levrini, 2008; Levrini et al., 2009).

According to the model, the longitudinal development of physics teaching/learning is schematically seen as a continuous widening of phenomenological evidence correlated to more and more powerful modelling along 3 macro-levels of development:

- Level one (K-8) - phenomenological "appropriation" of common life experience;

- Level two (grades 9-10, the first two years of upper secondary school) - re-arrangement of phenomenologies into pre-theory structures (nets of physical concepts);
- Level three (grades 11-13, the last three years of upper secondary school) - re-arrangement of the previously acquired knowledge in terms of physical theories, up to the basic concepts and formal insights of "modern" physics.

The structure of physical knowledge as discipline (i.e. its organization in terms of theories roughly corresponding to the traditional "chapters" in which physics is usually taught) is not taken in the model as the basis on which the whole long-term curriculum can be organized. Rather, the disciplinary structure of physical knowledge is considered to be a high cultural goal to which students will be progressively and consciously guided in the last years of upper secondary school.

In this study, MLD acted as framework for action. The experiment on teaching/learning thermodynamics, described in the next section, concerns the level Level 3 ($L_3$) and has been designed according to the following criteria, assumed within MLD as general orienting principles for the design of the experiment:

a. *The role of productive complexities in a physics reconstruction for all students (physics as culture).* The teaching/learning process is assumed to be meaningful if it tunes, in a productive way, three very complex systems: Real world, physics and cognitive dynamics. Because of the involved forms of complexities, hyper-simplified instructional descriptions and explanations, by making the material seem easy, can dangerously distort the learning process as well as the content. On the contrary, once useless complications are avoided, forms of unavoidable complexities can be exploited as productive for creating learning environments rich enough to enable each student to exploit his/her potential and to follow the route more congenial to his/her cognitive style and/or cultural interest. In the proposal on thermodynamics described below, the two main sources of productive complexities considered are: (1) The comparison between different approaches to the same content (a multi-perspective approach) and (2) the plurality of the dimensions tuned in the proposal (a multi-dimensional proposal). Such forms of complexities have been investigated by our research team for several years, in particular in studies about teaching/learning modern physics (special relativity and quantum physics) (see, for example, Levrini et al., 2008).

b. *The distinction between classroom corridors and students' individual trajectories.* The two terms "corridor" and "trajectory" are taken from the work of Confrey (2006) who asserts the necessity for research to develop teaching paths in actual classrooms in collaboration with researcher-teachers in order to reach two kind of goals: i) To provide teachers "of guidance (…) to organize corridors (…) as intellectual spaces through which students progress" realizing "a particular conceptual trajectory through the corridor", "with variations among students", and ii) To allow the researchers to identify the invariants inside the different paths and to use them in "the articulation of domain-specific guidance illuminating conceptual corridors" and in "engineering and identifying multiple means for constructing" corridors that can be successfully navigated in different ways by different classes.

In short, according to the MLD, the criteria used for designing the teaching materials and strategies are strongly oriented by the aim of fostering the creation of the learning environment as a *properly complex territory*, that is:

- able to outline a wide, coherent perspective where productive complexities are exploited (*corridor*);
- suitable to be "navigated" by the students along different *trajectories* according to different cognitive styles and/or cultural interests.

## 3. The study

### The context

The study concerns a teaching/learning experiment carried out during regular activities within a class of a science-oriented secondary school (Liceo Scientifico "A. Einstein") in Rimini, Italy. The class, composed of twenty 17-year-old students (9 girls; 11 boys), was attending the 4$^{th}$ year of upper secondary school (12$^{th}$ grade) and had studied physics from the first year (9$^{th}$ grade). Following the MLD, during the first two years (Level $L_2$), the teacher (PF, one the authors) carried out activities strongly focused on the problem of guiding students to construct pre-theory structures (nets of physical concepts) suitable to manage classes of phenomenologies (concerning motion, heating, light and vision, waves). The re-arrangement of the acquired knowledge in terms of physical theories ($L_3$) was the general aim of the activities carried out from the third year (11$^{th}$ grade). Before the teaching of thermodynamics as a theory, the curriculum had covered classical mechanics and special relativity. The thermodynamics corridor implemented in class was designed by a research group made up of researchers in PER, experienced teachers and undergraduate students.

According to the MLD criteria, the design of the thermodynamics corridor was characterized by the following basic choices:

- to present the same contents from different approaches. In this case, the two approaches were the macroscopic and microscopic ones, treated consistently as two different models, that can be compared by confronting the complementary explanations the two models give of the same phenomenology;

- to develop a critical-epistemological reflection on the peculiarities of the two approaches through specific activities (a questionnaire where students were asked to analyze and discuss excerpts of epistemological texts on the distinction between the two approaches, submitted at the beginning of the thermodynamics study; two collective discussions about the questionnaire performed at the beginning and at the end of the whole work);
- to emphasize the comparison of thermodynamics with the theories previously studied (classical mechanics and special relativity), so as to support learning as a continuous process of refining, re-investing, revising the models and the formal structures already acquired by looking at what was already known from new and more powerful perspectives.

The previous experiences of the group in designing teaching materials of modern physics (Levrini 2002) have provided interesting results about both students' understanding of the physical concepts (Levrini & diSessa, 2008), and the development of students' intellectual autonomy in properly complex territories (Levrini et al. 2008).

Following the research results of many years of investigation about teaching/learning modern physics, the design of the thermodynamics proposal was explicitly shaped by the following two research-based assumptions:

1) A multi-perspective approach can foster deep understanding of the physics concepts.
   In particular, the model of coordination class of learning and conceptual change (diSessa & Sherin, 1998), used as theoretical framework for analyzing a complex classroom episode, provided a theory-based explanation of *why* explicitly exposing, managing, and relating multiple classes of projections of a physical concept (the concept of proper time in special relativity), also elicited by the comparison between the geometrical and operational approach to special relativity, seems to be a good instructional technique to work around documented difficulties in conceptual change (in special relativity) (Levrini & diSessa, 2008);
2) The multiple dimensions (phenomenological, formal, conceptual, metacognitive) tuned by the analysis of historical-epistemological debates on the interpretation of the formalism of modern physics can foster the creation of a complex learning environment where the students are stimulated to reflect on the intricacies of the opinions and their personal involvement with them.
   A heuristic study on the implementation of a teaching proposal on quantum physics enabled us to point out that the multi-dimensional learning environment fostered students' intellectual autonomy by encouraging them to generate the questions that puzzled them and find out explicitly productive answers by themselves (Levrini et al., 2008).

The main methodological feature that distinguishes the research on thermodynamics is the focus on student's individual trajectories. The pieces of evidence about students' trajectories collected in our previous studies on different topics mainly came from classroom discussions and therefore the elements of intellectual autonomy emerged from the occasional individual contributions. In this study the aim of pursuing the individual trajectory was clear from the beginning and specific activities and instruments were designed for this purpose.

## Aims, data sources and methodology

The study here presented focuses on a selection of the data we collected during the whole experiment on teaching/learning thermodynamics. We focus on the transcripts of the individual semi-structured interviews that were conducted at the end of the whole work. The interviews were specifically performed for addressing some research questions that had arisen from the previous study on quantum physics: Are the properly complex territories defined above really suitable to be navigated by the students along their own learning trajectories in an explicit and conscious way? If so, is it possible to point out what identifies an individual trajectory with respect to the classroom corridor? What features of corridors, teaching strategies and/or collective activities can trigger the construction of personal trajectories?

The first two questions will be answered fairly precisely by means of the data analyzed so far, whilst as far as the third question is concerned, the study will provide a preliminary answer in terms of hypotheses about the possible factors that enabled students to construct their own trajectories.

Out of the 20 students of the class, in order to avoid interfering too much with the regular classroom activities, 10 students (5 boys and 5 girls) have been selected for interviews so as to represent students' variety as far as motivation to study physics, special interest in thermodynamics, contribution to classroom discussions, level of performance in regular physics tasks are concerned.

The interview protocol included:

- general questions like: *What did you find more and less interesting in the work you did on thermodynamics? Why? What did you find easier and what more difficult?*
- specific questions addressed to the individual student for clarifying something he/she had said during the classroom activities, like: *During the last discussion in class you mentioned the existence of an interesting relation between thermodynamics and the theories you previously studied (like mechanics). Would you mind telling us more about this?*

The interviews have been conducted by a masters student (GT) in the presence of a researcher (OL) and another masters student (MS). They have been audio-recorded and transcribed. In the first step of the heuristic study, we decided to select 5 out of the 10 interviews so as to analyze a reduced amount of data and to make the search for criteria to analyze the whole set of data easier (Glaser & Strauss, 1967). The analysis of the 5 interviews produced 5 profiles that are illustrated in the next sections. The analysis allowed us to better define the nature of the trajectories that characterize each student's way of navigating the classroom corridor.

## Results

Confrey's distinction between "corridors" and "trajectories" is strongly evocative in pointing out the different roles of the teacher and the students in the teaching/learning episodes. In the design phase of the study on thermodynamics, the distinction was taken as a synthetic way to emphasize the need both of assuring an active role to the students in constructing their own learning trajectories, and of dedicating research and teaching efforts in designing learning environments (corridors) able to foster a resonance between the growth of the individual intellectual autonomy and the collective classroom progression.

A major research product of the construction of the students' profiles was the tentative redefinition of the concept of trajectory that appears to enable us both to extract deeper information from data and to contribute to the development of a theoretical frame of analysis.

At the present level of the analysis, in order to share and discuss with other researchers the results of our attempt at better defining the concept of trajectory, we think it useful to report the 5 profiles reconstructed from the interviews with the students. For this purpose we also include whole original sentences from the students' protocols so as to respect to the extent possible, the students' style of argumentation.

### Michele: Actuality of real life and curiosity for "how things work"

Michele, at the end of his interview, summarizes very clearly his position: "*I like Physics because it explains <u>how reality works</u>, so to say, I'm very curious about <u>how objects work</u> and natural events…*".

Next year Michele will start Engineering at the university. His interests are already precise. The topics that were most impressive for him were the definition of temperature and the kinetic theory of gases that "*better explains the causes, that is what temperature is, the very word, whereas in the macro* [approach] *we study the effects without being able to give a satisfactory explanation of the word*", but even more the Carnot Cycle: "*real life is there, let's say, these processes, the engines working through these cycles, therefore I was more interested in connecting it to real life than in knowing exactly what it is*".

He had previously been highly interested in classical mechanics, getting very good marks on it: *"we were looking for causes there"* as he said, it was a *"framework"* for understanding *"things of the real world"*, like engines, he was very keen on. While studying thermodynamics he was very quick in dealing with formal steps for expressing ideal gas laws at microscopic level and in studying heat engines. His out-of-school interests and his personal objectives have driven also his school-work that gives him satisfaction.

Also because of his very precise motivation, Michele played an important role in the classroom dynamics by taking a very clear position both in favour of the microscopic approach to the first law and in favour of the macroscopic approach to the second law: He became a stable and respected reference, often opposed to Matteo's position.

In spite of his precise preferences, Michele expressed his interest in the whole teaching path, because "the two ways led to reality" and gave an example of how "scientists work":

"*It seemed to me very interesting for understanding…. these two different ways to manage to arrive to principles, to how reality was investigated. It was very interesting to me also for understanding how scientists work, how Einstein, Poincaré worked…"*

### Matteo: The pleasure of speculating and the search for approaching science from a "humanistic" point of view

The sentence that better summarizes Matteo's position is: "*I take for granted that <u>it is fundamental to speculate on reality and life</u>, because when you see phenomena and you get formulas out of that, then you have nothing, no foundation on which to stand. […] I believe that in order to find a way to understand a method to reach a reality […] it is fundamental to build a basis and <u>to speculate on how theories are found</u>, how concepts are elaborated. These concepts will certainly last longer than some formula."*

Matteo appears to like neither the formal elements of physics, "*too abstract*", nor the science claim to arrive at the truth: "*In my opinion Physics should not aim at truth, it should only […] be useful for mankind […]. To try to believe it possible to find in phenomena the foundation of truth is, I believe, an assumption out of the physics domain*".

The main point of his argument is that, in doing science, "human" and "far from reality" frameworks are projected on the world: *[…] <u>the sciences seem to me mainly human constructions</u>, I mean the way man has to investigate nature, to try to explain, may be without understanding that we cannot explain nature because we* *are human, we can only explain by the means we have… [and these] in my opinion are not powerful enough to catch the laws of nature… they are all human constructions that seem to me very inaccurate*".

Matteo played a role in the discussions similar to that of Michele, taking very precise and opposite positions to Michele's ones by criticizing the kinetic theory of gases and the microscopic approach in general: "*yes, the approach I prefer, you may already know, is the macroscopic one, because it stops at observing what happens and does not try to investigate reality, does not aim to understanding how these things happen from the inside, like the microscopic theory does by probing the bodies and finding some particles that move the same way… strange things*".

His opinion of science takes a positive turn when Matteo judges the epistemological activities carried out in the classroom. The topic he particularly enjoyed was the concept of irreversibility connected to the time arrow: "*yes, it is exactly this about entropy, where it develops a way of looking at the world which is a little less naive, because we see that the world is becoming, that some things that happen cannot be transformed… taken back to how they were: people get older, the table falls over, the cup breaks… cannot be put together again and here is a key concept to manage to accept even somehow reality, trying to explain everything following the path we went with these laws of mechanics, which is an odd idea, because mechanics laws are reversible and therefore there is no reason for a time line and for the time to go ahead; this may be a too philosophical way of looking at reality, because this way of refusing the becoming of being, of wishing to explain everything, the fact we understand that something cannot be explained, that there is this time advancing, this becoming that cannot be reduced to a being, may be this is something I enjoyed, I mean I enjoyed it*".

**Lorenzo: The pursuit of a unified consistent view of physics**

Lorenzo makes his position explicit by saying: "*One can see that everything is resumed, it is <u>not divided in topics</u> each with its own laws, instead everything can be <u>connected, unified</u>; the argument becomes <u>wider, more uniform</u>*".

Lorenzo appreciated the macroscopic construction of the ideal gas equation because of its unifying nature: "*It was rather quick, with those three laws [Boyle's and Gay Lussac's ones] we quickly arrived to the ideal gas law"* and its link to mechanics ("*trying to explain again temperature by referring to particle dynamics*). He confirms his position when talking about the comparison of the two approaches: "*that in the end I managed to understand which elements I prefer of one and of the other… I preferred one for some features and the other for some others, in the end I managed to make everything uniform, therefore I understood the whole argument rather clearly, thanks to these two ways*".

"Understanding" means, for Lorenzo, "connecting", "unify in a general framework". His focus of attention is the systematic pursuit of consistency: to solve contradictions, to get rid of spurious elements, by bringing everything back to the simplest possible framework (in his opinion mechanics is the simplest theory).

Lorenzo admitted to have met serious difficulties in completing the questionnaire at the beginning of the teaching path: the key was lacking for giving a unified structure to the lot of information contained in the excerpts. The comparison between the two different approaches proved very satisfactory to his needs: "*…when I read it again (the questionnaire) […] I immediately associated microscopic thermodynamics to the constructive type of theory, at once did it come to me….. and I read them (the excerpts) in an altogether different way, … and it was all clearer than the first time, I managed to connect and realized immediately that something had changed… […] and I felt satisfied… because I had then only to read… whereas the first time it had been hard, it took the whole afternoon, I had to read it many times, many passages were obscure….*".

He concludes by saying: "*[…] I noticed a big difference between reading it at the beginning and at the end, I mean that hadn't I read it at the beginning I might not realize all the work we have done*".

Lorenzo played indeed a fundamental role in the final discussion about the questionnaire: his need for consistency was a basic support to reflection and an incentive toward pointing out "contradictions" of the overall reasoning (the "jarring notes").

**Chiara B.: "Understanding as seeing" and the systematic effort of widening "one's own views" looking for and testing "new points of view"**

In the final discussion on the questionnaire Chiara said: "*We have analyzed thermodynamics from both perspectives, <u>looking</u> at differences and similarities, I mean we have two different <u>points of view</u> ... it may be more complete to try to analyze a phenomenon, or whatever is around us, from two different <u>points of view</u> rather than from one and therefore this teaching path widens what we've done. It widened <u>my view</u>*".

"To understand", for Chiara, means "to see". Her interview is full of this metaphor. The comparison between the two approaches, the two "points of view", as she says, takes a more general, social and cultural meaning. In learning physics, the comparison has been very interesting because it gave various opportunities to "play" between the real and the ideal and to use her metaphor of "seeing": from empirical observation of reality to the investigation through the microscopic approach, of what can only be seen with the "mind's eyes".

"*Talking of the concept of temperature... one* [approach] *is more tangible, more real, because you see it with a thermometer, you measure, you say «yes, there is a change in temperature, I see it, I understand it», but* *also the microscopic one is useful [...] The particle movement, for instance, I cannot see it in a thermometer, only going to the micro can I see this movement*".

Therefore the microscopic approach has been useful from a personal point of view: "*Not so much for better understanding the macro, that I found easier, but because it gave me a different point of view*".

Her attitude towards physics comes out also when she speaks about irreversibility. In the following sentence Chiara expresses her surprise in finding also in the concepts of irreversibility the overturn of reality that allows physics to go beyond appearance: "*[I liked] entropy because it was interesting this business of irreversibility and reversibility of phenomena, ... I told to myself «what a strange thing: phenomena are irreversible, there is a before and an after, but I take them for reversible and also here I make another ideal model, I consider them through other processes», I mean.... it's* <u>*some kind of overturn*</u> *always in physics and that's what I like...*"

Chiara plays an important role in both classroom discussions on the questionnaire because of her passion for "*seeing other people's points of view*", for "*widening one's own view and avoiding getting fixed on some issues*".

**Chiara C.: Focus on details and search for "not obvious details that are taken for granted"**

"*I like physics because I sometimes happen to ask myself... I mean I am rather curious about how things happen*". What Chiara seems to like the best of physics is that "obvious things are not taken for granted".

Of the thermodynamics corridor she enjoyed everything "<u>*that made me notice something I had not noticed before*</u>", like the meaning of measuring an apparently obvious quantity like temperature or the fact that "*a true theory is the one, like thermodynamics, that manages to give the same results from both approaches* [... I believe that] *a theory that really works is the one that manages to compare the two ways that can be verified*".

Chiara pays attention to the smallest details that she keeps under control to be sure that the constructed model of reality is not too idealized or unfounded. "*In the microscopic I liked and disliked at the same time this business of the model of a discrete gas, ideal, rarefied...*". What she dislikes of the model is the idea of "perfection" that appears to give a too ideal and simplified image of reality, neglecting too many details. What she likes is "*to be able to think to a model we can make experiments on and to create such a perfect model that everybody would be able to imagine a thing like that*".

Once she has been reassured that the model is rooted in reality, she can appreciate it for making her notice details she had not noticed before: "*I liked the macro best because* <u>*it made me think of things that I took for granted*</u>*, whereas I liked the microscopic because there is this model that makes you think, because if you focus on all the passages to arrive to the final law, we had to take into account a lot of things*".

Chiara played an important role in classroom dynamics because of her attention to details that, when recognized by all students, became catalysts of the discussions.

The idiosyncratic nature of the way each student reacted to the teaching proposal is evident from the excerpts as well as his/her ability to move within thermodynamics showing confidence and competence on the topic (appropriation).

What features of the profiles allow us to say that they constructed their own trajectories within the classroom corridor?

Looking again at the profiles in order to answer to this question we identified a set of criteria that we implicitly used in tracing the profiles and that could be used as an explicit tool for re-analyzing the whole set of data in terms of students' trajectories:

- the line of thought is *explicit* in the words of the student;
- the line of thought is *consistent* across the whole interview;
- the specific approach to disciplinary knowledge is *not occasional* (it can be traced back to the student reaction in various classroom activities);
- the arguments are *on-task*, referring to selected aspects already present in the classroom corridor;
- the keywords, around which the student constructs her/his arguments, reveal personal engagement, by being *genuine* and *emotion-laden*.

## 4. Final reflections and implications for research

The heuristic study here presented started from a "framework for action" (diSessa & Cobb, 2004) that we developed according to the MLD produced within the Italian project F21. Such a framework oriented us in designing and managing the learning environment as properly complex territory. As far as the research questions are concerned, the study presented here shows that the learning environment really proved to be able to foster students' intellectual autonomy. More specifically, the profiles outlined from 5 interviews allowed us to point out specific features that characterize the individual trajectories with respect to the classroom corridor. In this sense the results, although preliminary, represent an advancement in knowledge on design studies. The results moreover illuminate some foci of attention for analyzing the data and they point out the need for programmatic and iterative investigations over time along different directions:

1. *Re-shaping the framework for action in more detail* and designing future experiments on teaching/learning thermodynamics as "successive iterations that can play a role similar to systematic variation in experiment" (Confrey, 2006). At present, two new experiments are planned: Experiments where the same corridor and the same tasks are proposed to 17-year-old students by different teachers in a different school. This will be the opportunity for testing the effects that the new awareness about students trajectories can have: Their existence can indeed encourage teachers to foster their construction by the students but it can also affect the deeply genuine character of the trajectories described here. In any case, we expect the comparison of these variations of the same experiment to provide new hints for studying the classroom dynamics by cutting the process on the joint that lies between corridors' and trajectories' features.
2. *Analyzing trajectories within the more general problem of conceptual change* and, more precisely, within the problem of *a longitudinal approach to "appropriation" of science ideas*. Meaningful learning of a topic requires, in a quite obvious way, a long-term process of appropriation. By appropriation we mean the process leading to enter a topic deeply – and extensively – enough to see that topic in a personal version. In other studies about appropriation, concerning teachers coping with innovative research proposals (De Ambrosis & Levrini, submitted), it has been pointed out to what extent an appropriation process, that requires one to find one's own point of view, is a problematic task because of the complex dynamics of finding criteria for coherence between (critical) details and a global perspective (Viennot et al., 2005). The students' trajectories described above seem to have all the features of an appropriation process, because of the personal, genuine, character as well as the coherence they have in situating contents details in a global framework sensible to them. The specific point of how, why and under what conditions the construction of personal trajectories is connected to deep understanding (conceptual change) deserves future research studies, both on the data examined in this study (by strictly crossing the personal trajectories and the contents comprehension), and on other data where, for example, the personal trajectories seem to distort the learning of a concept and where the process of appropriation and/or conceptual changes seem to be problematic and/or unsuccessful.
3. *Individuating the factors that can trigger intellectual autonomy* by the construction of personal learning trajectories. A study on "personal excursions"[1], in spite of their different features with respect to students' trajectories, inspired us in the search for those factors that can made the trajectories' construction possible (Azevedo, 2006). In agreement with Azevedo, we think that such factors need not to be identified at a too general level: They have indeed "to operate at a level of specificity that allows for a satisfactory explanation of the dynamics patterns characteristic of personal excursions [personal trajectories, in our case]". Following his suggestions, we think that, as a preliminary answer to the third research question ("What features of corridors, teaching strategies and/or collective activities can trigger the construction of personal trajectories?"), the following factors can be identified:
    - *the specific feature of the content reconstruction* that, by means of the comparison between two different approaches (macro and micro), is able to create a collective dynamics where students are involved with different roles (as partisans, opponents, moderators, disrupters, facilitators, …). Such a dynamics seems to have a productive power in making each student to find her/his more congenial path and progressively to test and refine it.
    - *the features of the activities specifically addressed to activate an epistemological meta-discussion* of the peculiarities of the two approaches. Although unusual, unfamiliar and demanding, the design of such activities has been always carried out in such a way that students could find themselves in their "regime of competence" (diSessa, 2000, quoted in Azevedo, 2006): Activities where they could approach the new epistemological language by applying it to physics domains previously studied.
    - *the choice of the teacher*, made explicit to the students, *to evaluate* their performance in the epistemological activities, by taking into account not "what they wrote or said in specific", but *their level of involvement*, the *coherence of their thinking* and the *contribution they are able to give in the collective discussions*.

All these factors deserve further studies specifically addressed to improve our theoretical understanding of students' trajectories. So far they seem to suggest that students' trajectories are stimulated and become evident in contexts where the whole teaching proposal is systematically developed and managed by the teacher through actions directed to move the learning problems from a plan where only “right-wrong”, “yes-no”, “bad-good” answers are possible, to a plan where the problems themselves, although respecting the shared disciplinary knowledge, admit an analysis in terms of critical evaluation of different possible, consistent, “points of view”.

## Endnotes

(1) Personal excursions are defined by Azevedo as self-initiated, self-directed, self-motivated and relatively long-term activities, belonging to a student’s trajectory but task-off with respect to the classroom corridor. They represent “key events whereby students connect, in a deep and personal way, to the subject matter and overarching goals [of the unit proposed to them]” (Azevedo, 2006).

## References

Azevedo F. S., (2006). Personal Excursions: Investigating the Dynamics of Student Engagement. *International Journal of Computers for Mathematical Learning,* 11:57-98.

Confrey J., (2006). The evolution of design studies as methodology, in Sawyer K. (ed.), The *Cambridge Handbook of The Learning Sciences*, Cambridge University Press, 135 – 152.

De Ambrosis A., Levrini O., (submitted), How physics teachers cope with innovative teaching proposals: An empirical study for reconstructing the appropriation path in the case of Special Relativity.

diSessa, A. A., Sherin B. L. (1998). What changes in conceptual change? *International Journal of Science Education, 20*(10), 1155–1191.

diSessa, A. A., Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *The Journal of Learning Sciences,* 13(1), 77-103.

Glaser, B. G., Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Guidoni P., Levrini O. (eds.) (2008), *Approcci e proposte per l'insegnamento-apprendimento della fisica a livello preuniversitario, dal progetto PRIN F 21*, Forum Editrice, Udine.

Kattmann U., Duit R., Gropengießer H., Komorek M. (1996). Educational Reconstruction -bringing together issues of scientific clarification and students' conceptions-, paper presented at the Annual Meeting of National Association of Research in Science Teaching (NARST), St. Louis (MI), April 1996.

Levrini O., (2002). The substantivalist view of spacetime proposed by Minkowski and its educational implications, *Science & Education*, 11, 6, 601-617.

Levrini, O., diSessa, A A., (2008). How Students Learn from Multiple Contexts and Definitions: Proper Time as a Coordination Class. *Physical Review - Special Topics Physics Education Research*, 4, 010107-1-18 (2008).

Levrini O., Fantini P., Pecori B. (2008), “The problem is not understanding the theory, but accepting it”: a study on students’ difficulties in coping with Quantum Physics, in R. Jurdana-Sepic R., V. Labinac, M. Zuvic-Butorac, A. Susac (Eds.), *GIREP-EPEC Conference, Frontiers of Physics Education (2007, Opatija), Selected contributions*, 319-324, Rijeka: Zlatni rez (ISBN 978-953-55066-1-4).

Levrini O., Altamore A., Balzano E., Bertozzi E., Gagliardi M., Giordano E., Guidoni P., Rinaudo G., Tarsitani C. (2009), Looking at the physics curriculum in terms of framing ideas, proceedings GIREP 2008 International Conference, MPTL 13th Workshop, Nicosia, Cyprus, August 18-22, 2008.

Viennot L., Chauvet F., Colin P., Rebmann G., (2005). Designing Strategies and Tools for Teacher Training: The Role of Critical Details, Examples in Optics*, Science Education*, 89 (1), 13-27.

# Designing Assessments to Track Student Progress

Namsoo Shin, Shawn Y. Stevens, & Joseph Krajcik
University of Michigan, Ann Arbor, MI
Email: namsoo@umich. edu, sstevens @umich. edu, krajcik@umich. edu

Learning scientists have embraced the idea of using learning progressions for the development of assessments to monitor learning over time. Using LPs to develop assessments requires an iterative, process-oriented approach, and involves design products that work in real contexts. This proposal illustrates a design research process, Construct-Centered Design, for the development of assessment tasks to track the long-term development of student learning for a core idea in science – the transformation of matter. Fifty-six items of various types were designed to measure student understanding along the lower levels of a LP. We iteratively collected cross-sectional data from 500 middle students for developing, piloting, and revising the assessment items. We conclude the paper by discussing the strengths of and weaknesses of this process for measuring students' understanding across levels of the LP. The ultimate value of this work will propose the design process of assessment to track student learning.

Learning scientists have emphasized the importance of intensive research for addressing theoretical questions about the nature of learning in context and producing *evidence-based claims* to support their theoretical questions (Collins, Joseph, & Bielaczcy, 2004). Moreover, for valid evidence-based claims, researchers need to consider student learning over an extended period of time because learning challenging content takes years to develop (Stevens, Delgado, & Krajcik, 2009; Duschl, Schweingruber, & Shouse, 2007). Such research requires the development of assessment tasks (e.g., interview protocol, test item, observation instrument, survey) that can track student progress over time in real contexts rather than simply examining isolated variables within laboratory contexts (Brown, 1992; Barab & Squire, 2004) or memorized fragmented pieces of knowledge. However, to develop such assessment tasks is challenging. The learning science community needs models that illustrate principled and systematic ways of doing so. Recently, the science education and learning science communities have embraced the idea of learning progressions (LP) to provide a guide for monitoring student learning over time (see for example, Smith et al., 2006).

Learning progressions are research-based descriptions of how students may build their knowledge, and gain more expertise within and across a core idea over a broad span of time (Duschl, Schweingruber & Shouse, 2007; Smith et al., 2006). They illuminate how learners can develop and connect concepts within and across disciplines as they progress towards a more sophisticated understanding of the key concepts and skills necessary. As such, LPs can provide a potential path for students to develop understanding of core ideas over time, and can guide the alignment of instructional materials, instruction and assessment in a principled way to support the development of integrated learning (Duschl, Schweingruber, & Shouse, 2007; NRC, 2006 & 2007; National Assessment Governing Board, 2006a & 2006b). A LP contains three key factors: a lower and an upper anchor to define the range of content within a core idea and defined levels of understanding between the lower and upper anchors (Smith et al., 2006; Stevens, Delgado, & Krajcik, accepted). The levels of a LP specify not only the order in which students develop understanding of the important concepts, but also how they *interconnect* and *reason with* the important concepts between related ideas.

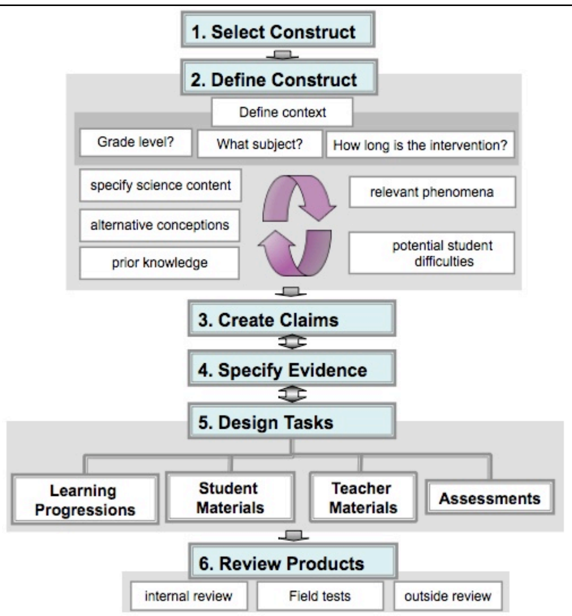
However, developing assessments that align with a LP requires an iterative, process-oriented approach, and involves designing products that work in real contexts. Because there are no such ready-made LPs, an iterative process of building, validating, and revising LPs and associated assessments using exemplary instructional materials is critical. Because learning is a complex process, many factors affect the path that students may follow as they build understanding, including the learning context, instructional materials, instruction, and students' prior knowledge and experiences. Thus, in order to build LPs and associated assessments, empirical data need to be collected from students who have experienced curriculum materials that were developed following LPs and learning principles. This helps us to ensure that students' lack of understanding is not because of inadequate learning experiences, but because of the developmentally challenging ideas students are expected to learn. Well-developed coherent curriculum materials and associated assessments based on a LP should be designed, implemented, and tested iteratively throughout the process of refining a LP. In addition, intensive research is needed to better characterize student understanding because gaps in learning research still exists. These gaps need to be filled prior to building LPs. In sum, developing assessment tasks using LPs is a complicated process and requires thorough, longitudinal studies related to how students learn core ideas over time in diverse contexts. A principled research design process should guide the complex, iterative process of developing assessments along an LP.

In this paper, we illustrate a research design process that can be used to develop assessment tasks to track the long-term development of student learning for a core idea in science – the transformation of matter.

We explore the following research question: *How can assessment tasks be developed that monitor student learning over time and that assign where student understanding lies along the LP?* Based on our previous work, we propose using Construct-Centered Design (CCD) (Shin, Stevens, & Krajcik, 2010; Pellegrino, et al., 2008) to develop assessments based on LPs. Because CCD focuses on the construct that students are expected to learn as well as what researchers and teachers want to measure, the CCD process provides a flexible and systematic approach for guiding product development, monitoring the development process, and examining the effects on learning outcomes. Next, we describe the foundation of CCD and present how the CCD process applies to the development of assessment to track student understanding in a LP. Finally, we conclude by discussing the strengths of and weaknesses of the process.

## Research Design Process: Construct-Centered Design

A research design process to guide learning research and the development of products should provide flexibility for (a) mapping out the constructs associated with core ideas and (b) developing assessments and instructional materials that support and measure how student understanding develops over time. Thus, a design model should emphasize both defining constructs for instructional material development and specifying evidence for assessment development. By modifying and adapting the learning-goal-driven design (LGD) process for developing construct-focused curriculum materials (Krajcik, McNeill, & Reiser, 2008) and the evidence-centered design (ECD) model for developing assessments (Mislevy & Riconscente, 2005), the CCD process provides such a model. The CCD process begins with specifically defining the focus of the construct. We define the construct as the core ideas that students are expected to learn and researchers and teachers want to measure (Messick, 1994; Wilson, 2005). Because the foundation of the process focuses on the description and explicit specification of content that lies within constructs, the process is termed construct-centered design (CCD). In describing the process, we do not mean to imply this is a linear process. In practice, the process is interactive and highly recursive, with information specified at one step clarifying and often modifying what was specified earlier. Figure 1 illustrates the CCD process. A detailed description follows.




<u>Claim:</u> Student should be able to distinguish between matter and non-matter and justify their conclusions

<u>Evidence</u>: Students' evaluations and justifications should include the following criteria:

- Matter takes up space and has mass (has weight in a gravitational field)
- Matter can exist in three different states–solid, liquid and gas. The same substance can exist in all three forms under specific conditions

<u>Assessment task (multiple True or False item)</u>

Do you consider the following to be matter?

| | Yes | Not sure | No |
|---|---|---|---|
| a rock | | | |
| an idea | | | |

(…grain of salt, heat, a wish, air, a tree, a shadow, particle of chalk dust, beam of light, image in a mirror, smoke, helium inside a balloon, a cloud)

Which criteria are true of all matter?

| **Criteria** | **True** | **Not Sure** | **False** |
|---|---|---|---|
| Whether you can see it | | | |
| Whether you can touch it | | | |
| Whether it takes up space | | | |
| Whether it is alive | | | |
| Whether it is a solid | | | |

<u>Figure 1.</u> CCD process, Claim, Evidence, and multiple True or False Item example

### Select and Define the Construct

The first step in CCD is to choose the construct and define the target learners (see Figure 1, step 1). The construct is essential as it identifies the set of ideas for which learners will study and develop understanding. Because students at different grade ranges have different knowledge and experiences that influence their learning, defining the target students helps define the construct appropriately, and also guides preparation of level-appropriate instructional materials, instruction and assessment. The construct of our LP is the transformation of matter and includes core ideas associate with the atomic model and the interactions between atoms and molecules as they undergo various transformations. To help define the range of content that needs to be unpacked, the lower and upper anchors for the LP were defined. In this case, the lower anchor was defined using the learning progression for atomic molecular theory for grades K-8 (Smith, et al., 2006), and additional empirical research. The upper anchor of the LP was defined based upon national standards (AAAS, 1993; NRC, 1996), ideas required as a foundation for nanoscale science and engineering learning for grade 7-12 students (Stevens, Sutherland, & Krajcik, 2009) and current learning research related to those ideas.

The next step is to define the construct based on expert knowledge of the discipline and related learning research (see Figure 1, step 2). This process, called unpacking, involves defining the ideas contained within the construct. By unpacking, we mean breaking up the construct into smaller components to explicitly specifying the concepts that are crucial for developing an understanding of the construct. Being related to the construct is not enough; the concept must be necessary for building understanding of the construct. The depth of understanding that is expected from students is also clearly defined in this step. As a step towards defining *how* students should know the content, the prior knowledge that is required both within and from other related constructs is also specified. The unpacking process also includes: identifying potential difficulties students might have learning the content; specifying and clarifying non-normative ideas that might interfere with students learning the content; providing possible phenomena that may help student learn ideas and develop their understanding; and identifying strategies for effectively representing the concepts based on previous learning research. The concepts within the constructs were then unpacked to define what it means to understand them at levels appropriate for grade 7-14 students.

<u>Characterizing How Students Develop Understanding</u>

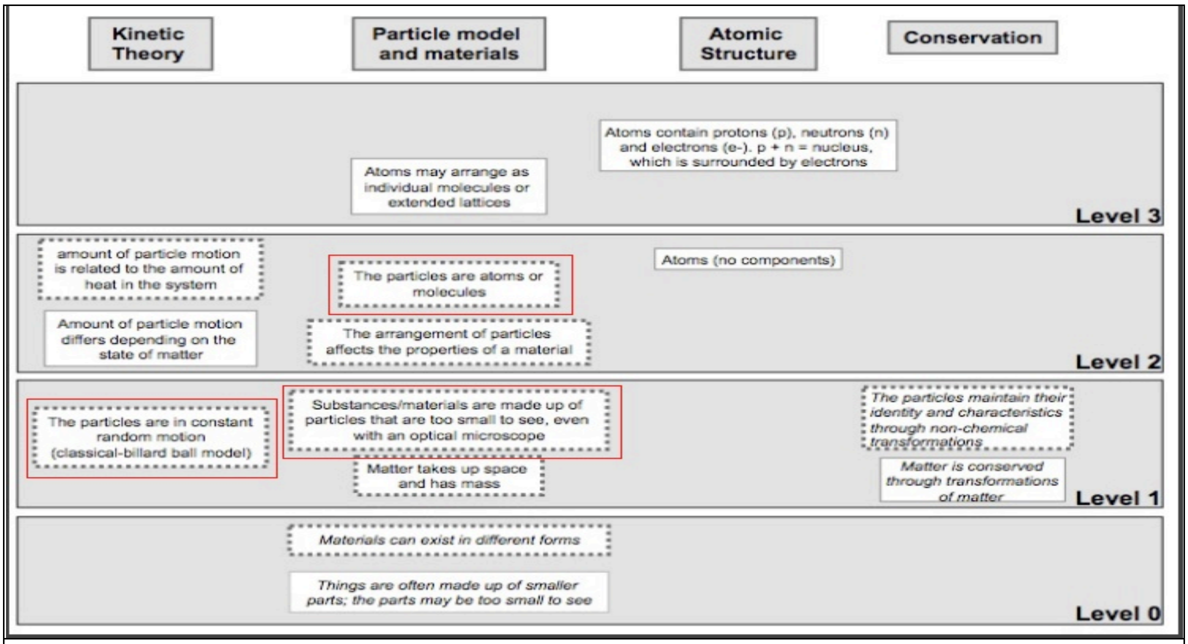
In the process of unpacking the construct, the current evidence-based learning research was not sufficient to completely define the levels of student understanding, possible non-normative ideas, and difficulties related to the developing understanding of the construct. To help fill gaps in the learning research related to the construct, an interview protocol was developed based on the CCD claim and evidence steps to characterize how ideas related to the transformation of matter developed as grade 7-14 students passed through the current curriculum (see claim and evidence sections for further details). A cross-sectional sample of students (N=79) representing the range of grades in our target population was interviewed. In this case, seventh grade students and high school students from the same district or school system were interviewed individually. At each level, we chose students who provided a mix of gender and a range of achievement levels. A 20-30 minute semi-structured interview was conducted with each student to characterize understanding related to sub-constructs of the transformation of matter. The interview questions required students to apply their knowledge to explain real world phenomena. The interviews were conducted in several phases. After each phase, student responses were evaluated and the protocol was revised to better characterize student understanding of the construct. The results of this study inform the strategies that may help students move along the LP. They provide information about how students develop understanding of the construct, and where they have incomplete knowledge. We used the results to describe prior knowledge about student learning using current instructional materials, and identify difficulties and non-normative ideas students hold regarding the content. Understanding students' ideas is critical in determining when and how it might be appropriate to introduce the concepts to students. In addition, knowledge of non-normative ideas aids the development of assessments that measure students' progress (see Stevens, Delgado & Krajcik, 2009 for detailed results).

## Create Claim(s) and Specify Evidence

A set of claims is generated based upon the unpacked construct and student data. Claims specify the nature of knowledge and understanding expected of students regarding a particular construct (see Figure 1, step 3). In constructing a claim, vague terms like *to know* and *to understand* are not used. Rather, claims specifically define what cognitive activities students will need to apply to respond to the task (e.g., Bloom's taxonomy; Bloom, 1956). For example, students should be able to provide examples of phenomena, explain patterns in data, construct explanations, or develop and test hypotheses. An important part of student learning involves the ability to connect ideas and apply knowledge to new situations (Bransford, Brown & Cocking, 1999). Therefore, it is important that the claims specify how students should connect ideas both within individual sub-constructs, and among related sub-constructs in order to describe how students build integrated understanding of the construct. The evidence specifies the aspects of student work (e.g., behaviors, performances) that would be indicative that a student has the desired knowledge to support a specific claim or set of claims (see Figure 1, step 4). In particular, this step explicitly defines the expected level and depth of understanding that the target learners should demonstrate. The understanding defined by the evidence provides a guide for specifying levels in the LP.

## Design Tasks

The tasks, which are generated based on the claims and evidence, provide a response that offers appropriate evidence to support the relevant claim (see Figure 1, step 5). The tasks can be either learning products that will help learners develop the knowledge in the claim, or assessment products that measure whether learners hold the knowledge stated in the claim. The assessment tasks are designed to elicit or generate students' performances to allow for a judgment to be made about whether sufficient evidence exists to support the learning *claim*. A single assessment task may provide evidence for more than one claim; multiple tasks may be necessary to assess a single claim. A single task or set of tasks can be associated with a claim or a set of claims assigned to multiple levels on the LP. An individual claim, its evidence and corresponding task may link to a single level on the progression. The right column in Figure 1 provides an example of a claim, its corresponding evidence and assessment tasks. Based upon the claims and evidence, we developed the first version of a LP (see Figure 2 for a portion of the LP).




**Likert-type item: The following are possible characteristics of atoms and molecules. Mark whether the statements are true or false.**

| Statement | T | PT | NS | PNT | NT |
|---|---|---|---|---|---|
| 1. At 75°F (~24°C) molecules and atoms are always moving. | | | | | |
| 2. Molecules are all the same size. | | | | | |
| 3. Molecules that make up ice cream are very cold. | | | | | |
| . | | | | | |
| 21. A balloon is filled with water. When squeezed the balloon changes shape. The pressure from squeezing also changed the shape of the water molecules. | | | | | |
| 22. The atoms in solid copper metal are hard. | | | | | |

T: true, PT: pretty sure it is true, NS: not sure, PNT: pretty sure it is not true, and NT: not true

Figure 2. A portion of LP and Likert-type item for measuring Level 1 and 2 of the LP

## Develop Assessment Items to Measure Levels of Understanding in a Learning Progression

The claims and evidence specify how students should be able to connect ideas both within a sub-construct, and among related sub-constructs in order to describe how students build integrated understanding. The claims and evidence can be used to refine the levels. Figure 2 illustrates part of the LP for the transformation of matter. In figure 2, the dashed boxes represent the content for which we are developing assessment tasks. The red color boxes correspond to the assessment item in the figure. The levels in the LP represent a set of ideas that describe a path towards developing a more complex understanding of the construct. The set of ideas within a level connects to explain a range of phenomena; higher levels describe the phenomena with greater scientific sophistication and completeness. In this way, the levels of the LP describe increasing levels of understanding. Further research is required to examine whether the order in which students learn concepts within a single level is important.

Based on the LP, we developed assessment items to locate the level of student understanding in the LP. Because the development of assessment items is a challenging activity, our design process will take several iterations to develop assessment tasks that will track student from the lower anchor to the upper anchor. In the first iteration, we focused on the development of assessment items for the lower level including levels 1 and 2. When possible, we selected previously published items that supported claims and evidence appropriate for the target students and modified them as necessary; otherwise, we designed and developed new items. We evaluated the items using the criteria of sufficiency, necessity and age appropriateness (Deboer, et al., 2008). We then revised the item stems, answer choices, and associated representations to ensure they support the claims and evidence associated with the lower levels of the LP and to ensure that text and representations were level appropriate. Fifty-six items using various item types (e.g., two-tiered, multiple true and false, complex multiple-choice, and Likert-type scale) were designed to measure how well students apply ideas within and across the sub-constructs. Although in future iteration, we will develop open-ended items, in the first Iteration, various item types as listed above were employed to develop assessment items that measure carefully and efficiently student understanding across grades 7-16 (Scalise, & Gifford, 2006) (see Figure 1, 2 and 3 for example).

## Review Products

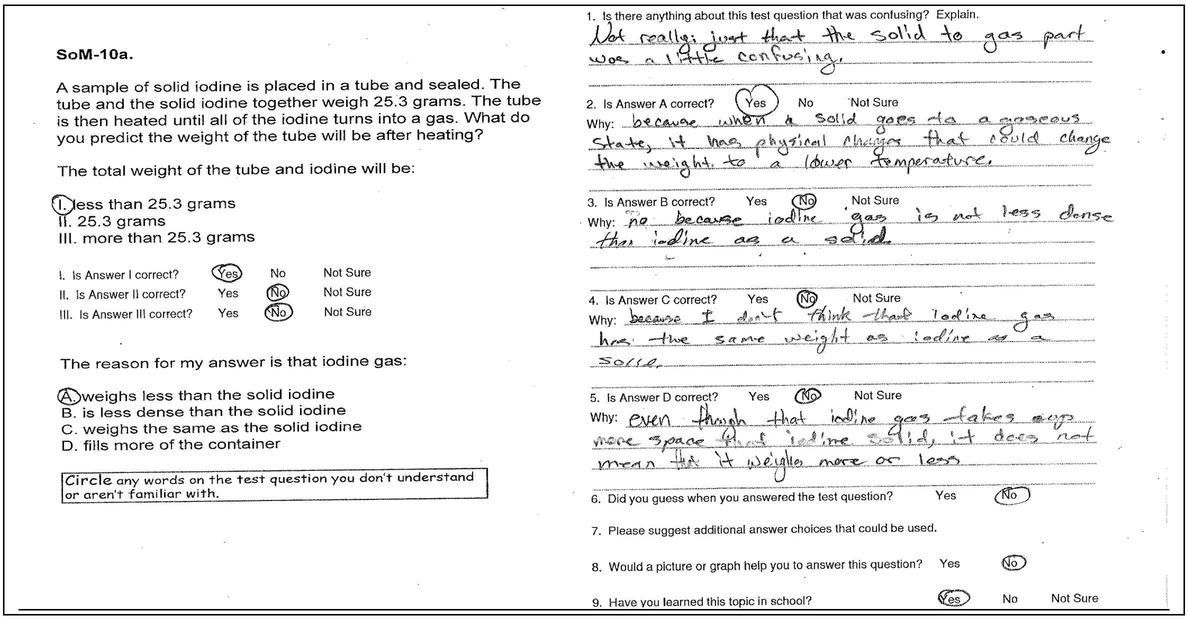
The next CCD step is to review the products. For each step within this iterative process, the products are reviewed internally and when appropriate, externally (see Figure 1, step 6). The internal review focuses on critique and revision of the products to ensure that they align with the claims and evidence. External review includes feedback from teachers of the target students or from content or assessment experts. Conducting pilot tests and field trials with target students is an essential component that provides invaluable information about the products.

### Pilot Study

In order to ensure that students interpret the questions and associated representations as intended, we piloted the items with 479 middle school students (grades 6-8) from three schools in two distinct communities representing a range of race, ethnicity and SES. The cross-sectional design ensures that items measure the understanding of students with a range of knowledge and abilities. A range of different types of assessment items can provide information about students' understanding. The different types of items vary in complexity (e.g., matching < categorizing < ranking and sequencing < assembling proof) and in the amount in which learners' responses are limited or *constrained* (e.g., traditional multiple choice vs. open-ended essay; Scalise & Gifford, 2006). Items that are more constrained are easier to score, but likely provide less information about student understanding and higher-order reasoning. In this pilot study, we investigated "*what are the advantages and limitations of different item types for measuring student understanding across Levels 1 and 2 of the LP?*" Each student was given a test form containing four to five items. We piloted each test item with 30-35 students. The piloted items were each accompanied by a set of questions to explore how students interpreted the item (Deboer, et al., 2008). Figure 3 illustrates the questions that accompanied the two-tier multiple-choice questions; slightly different sets of questions accompanied different item types.

We analyzed each item using a simple descriptive analysis, a classical item analysis and an Item Characteristic Curve analysis that focused on identifying item clarity. For the descriptive analysis, we created five categories: confusing words in an item, confusing item stem, reasons for choosing their answer, percent who correctly answered an item, and helpfulness of representation. We carefully reviewed the items that students answered correctly more than 80% and approximately less than 30% in order to investigate whether the items are too easy or difficult for target students, or are written poorly. Figure 3 provides an example of student responses to an item. In this case, students answered correctly 24% in the first-tier question and 33% in the second-tier question. This item posed conceptual and literacy difficulties for students. Because the item involved not only the idea of conservation of matter but sublimation as well, we judged that it might be a level 3 item, as students did not seem to understand the phenomenon of sublimation described in the item. Other students did not know that "iodine" could be a solid. In addition to ensuring that our pool of items were written clearly and at an appropriate level, student responses helped ensure that there is only one correct answer and provided an evaluation of the distractors. After the initial descriptive analysis, we then conducted an additional review of the two-tier multiple-choice and multiple-choice items using classical item analysis and Item Characteristic Curve (ICC). For the classical item analysis, we focused on the correlation between the individual item score and the total score of the test. Then, we reviewed the ICC of each item to investigate how well an item differentiates between students having ability below and above the item location using dichotomous data (1=correct, 0=incorrect), and how students respond to the distractors of the items using polytomous data (raw responses = 1, 2, 3, 4). These analyses helped us to revise each item stem, as well as the distractors for each item. For example, the results indicate that the two-tier item type was not appropriate for measuring middle-school students understanding because the students were confused about how they should answer the question, and the item did not provide additional information than when only using the second part of the two-tier question. We speculate that middle school students are not developmentally ready to interpret the structure of such an item type. We decided to conduct a future investigation for this item type because our finding is inconsistent with Treagust's study using high-school students (Treagust & Chandrasegaran, 2007)

We conducted an additional analysis to investigate the characteristics of the item types on measuring student understanding along the LP. First, we developed items using two different item types (multiple-choice and multiple True and False (T/F)) to compare how students perform differently on the two item types and whether the two item types provide similar or different information. We assessed student understanding using a Multiple T/F list as opposed to a series of multiple-choice questions for gaining insight into connections and/or the complexity of student understanding (see Figure 2). The results show that most high-ability students responded that the multiple T/F item type is more difficult than multiple-choice items. They reported that they have to think carefully before responding to the item because they need to respond to every answer rather than to select a single correct answer. Therefore, the multiple T/F item type may better assess student understanding because they cannot rely on test-taking skills such as answering a multiple-choice item. In contrast, a majority of low-ability students responded that the multiple T/F is easier than the multiple-choice item because they is not just one correct answer.


SoM-10a.

A sample of solid iodine is placed in a tube and sealed. The tube and the solid iodine together weigh 25.3 grams. The tube is then heated until all of the iodine turns into a gas. What do you predict the weight of the tube will be after heating?

The total weight of the tube and iodine will be:

I. less than 25.3 grams
II. 25.3 grams
III. more than 25.3 grams

I. Is Answer I correct? Yes No Not Sure
II. Is Answer II correct? Yes No Not Sure
III. Is Answer III correct? Yes No Not Sure

The reason for my answer is that iodine gas:

A. weighs less than the solid iodine
B. is less dense than the solid iodine
C. weighs the same as the solid iodine
D. fills more of the container

Circle any words on the test question you don't understand or aren't familiar with.

1. Is there anything about this test question that was confusing? Explain.
Not really; just that the solid to gas part was a little confusing.

2. Is Answer A correct? Yes No Not Sure
Why: because when a solid goes to a gaseous state, it has physical change that could change the weight to a lower temperature.

3. Is Answer B correct? Yes No Not Sure
Why: no because iodine gas is not less dense than iodine as a solid.

4. Is Answer C correct? Yes No Not Sure
Why: because I don't think that iodine gas has the same weight as iodine as a solid.

5. Is Answer D correct? Yes No Not Sure
Why: even though that iodine gas takes up more space than iodine solid, it does not mean that it weighs more or less.

6. Did you guess when you answered the test question? Yes No

7. Please suggest additional answer choices that could be used.

8. Would a picture or graph help you to answer this question? Yes No

9. Have you learned this topic in school? Yes No Not Sure

Figure 3. Two-tier type item and accompanying questionnaire with student response.

Second, we analyzed the relative difficulty of the Likert-type and multiple T/F items to explore how they measure the range of understanding for each level of the LP using Item Response Theory (IRT). We used a series of 14 T/F questions to measure how well students distinguish between matter and non-matter (See Figure 1). We treated each statement as a separate individual item of the multiple T/F item for the analysis. We reviewed the difficulty level of the items by examining a Wright map (Wilson, 2005) and an item fit output to ensure the items measure across Level 1 and 2 of the LP. The Wright map showed that item difficulty ranged between logit -2 and 1 out of a range from – 6 to 6. The – 2 to 1 range is consistent with our intention of measuring Level 1 and 2. A logit takes into consideration both the ability of respondents as well as the item difficulty in assigning a task to a particular location on the Wright map. In addition, we found that virtually all students classified the solid forms of matter–rock, a grain of salt, a tree and a particle of chalk dust–as matter. The item difficulty levels are very similar between logit -1 and 0. In the revision of the item, we streamlined the list by removing three of these objects since they were not providing additional information. The other part of the question was designed to gain information on students' reasoning as they distinguish between matter and non-matter. Students were asked about the criteria necessary for something to be classified as matter. Students in the first pilot suggested the "whether it's a solid" as a possible criteria. We found in subsequent administration of the item that it proved to be an effective distractor. We created a Likert-type item that surveys a range of ideas related to the atomic and kinetic theories based on a survey created by Harrison and colleagues (Harrison, Kease, & Voss, 2006) to efficiently measure students' models of the structure of matter (see Figure 2). Their difficulty levels are between logit -1.5 and 1.5, which is the range for Levels 1 and 2. Twenty-one statements were spread evenly across three logits on the Wright map. From the findings, we conclude that the Likert-type and multiple T/F type items can be used to measure a range of student understanding along the LP without spending significant testing administration time or giving up probing student reasoning.

## Interview Data

After data analysis, we selected problematic items where the difficulty was not clear (e.g., whether the problem resulted from students' lack of understanding or the item itself). The five selected items included a multiple-choice item with a picture, a two-tier item, a Likert-type item, a graphic representation item, and a model representation item. We developed a semi-structured interview focusing on the selected items and interviewed 19 middle school students to ensure the accompanying items adequately measured students' understanding. We analyzed the student data using the five categories described above.

In the case of the two-tier item, the interview confirmed that most students did not know what was "iodine." Their responses were similar to students in the pilot study. However, they could guess that it is a chemical substance and that it can be liquid and solid. We initially decided to use "iodine" in the question because the students could answer the question correctly without knowing is the meaning of "iodine." In addition, the interview results about the two-tier item confirmed that middle school students have difficulty understanding the structure of the item. They had to spend time trying to understand how to answer the item rather than how to apply their knowledge to answer it. From the pilot and interview data, we conclude that we will not employ the two-tier item type for future development of assessment items for middle school students. For the Likert-type item, students felt comfortable responding to the item, but felt it was more difficult than a multiple-choice item because they had to think about the degree of the correctness of each statement. Based on the pilot and interview data, the results confirmed that the Likert-type item is appropriate for measuring middle school students' level of understanding. The results from the graphic and model representation items provided detailed information about student's interpretations of the item stems, distractors' representations, and graphical representations to guide revision of the items for the next iteration. Overall, the results from the survey instrument and interview data were consistent, which means that in the future, we can use a survey instrument to collect data about items instead of collecting intensive, time-consuming qualitative data.

From the first iteration of data collection and analysis, and further expert and internal review, we are revising 11 problematic items for the next iteration. The important task of the next iteration is to analyze the relative difficulty of the items to gain insight into the relationship between the students and the assessment items using Item Response Theory (IRT). We will collect a student data set, approximately 100 students per item to analyze the relative difficulty of the items. This item analysis will confirm that the items measure student understanding corresponding to Levels 1 and 2 of the LP.

## Conclusion

We have described the iterative CCD research design process that researchers can use to develop assessment items that align with a LP to track learning across time. These items will provide a scale that can locate students on the lower levels of the LP. Ultimately, we will use these items to collect longitudinal data to evaluate student learning as they experience three years of coherent instruction that supports learning of the nature of matter.

A number of researchers have discussed the value of developing learning progressions to track students learning (Wilson, & Berenthal, 2006; Smith et al., 2006; Duschl, Schweingruber, & Shouse, 2007). However, there are still critical challenges to overcome in the process of developing and refining LPs and valid associated assessments to measure the level of student understanding across time (Pellegrino, Chudowsky, & Glaser, 2001; Smith, et al., 2006). First, because of the complexity of building LPs, we need to use a research design process that is iterative, process-oriented, and involves designing products that work in real contexts that extend our understandings of the nature and condition of learning and development as well as promote student learning (Barab & Squire, 2004; Brown & Collins, 1992; Collins, Joseph, & Bielaczcy, 2004). However, the typical strategy for this type of learning research employs naturalistic methodologies to investigate how learning occurs and the product development process for building evidence-based claims (Barab & Squire, 2004). A fundamental challenge for such research is the extensive quantity of qualitative and quantitative data that must be collected and organized in order to provide appropriate evidence to support the research claims (Collins, Joseph, & Bielaczcy, 2004). Second, although LPs offer a promising framework, limited validated assessment items exist which connect assessment items and the developmental progress of student understanding to illustrate conceptual growth. Prior to using the LP to track student progress, extensive research is needed to validate assessment items.

As we have illustrated, CCD can be a valuable process for learning research that may overcome these challenges. In particular, the CCD approach focuses on clearly defining the construct to focus the research and development strategies. Another critical characteristic of CCD is the explicitly specified evidence based on the unpacking of the construct that links directly to the claims. Specifying the claims and evidence supports the development and alignment of a range of products. The systematic process outlined by CCD provides guidance for the collection, organization and analysis of data by defining what data is essential for supporting the learning claims we hope to make about student learning. CCD can be considered a component of an iterative development process that is constantly being refined and revised to accommodate the needs of learning researchers. We still have much work to accomplish to make CCD a usable design model for other researchers. We need to further develop the guidelines and examples for each step of CCD to provide guidance on how researchers can use CCD to accomplish a variety of design-based research goals. To do this, researchers need to apply the CCD process to design various research, instructional materials, learning progressions and assessments tasks in order to articulate the subcomponents of the various CCD steps more clearly. As we and other researcher use CCD to guide a greater amount of research and development products, the process will become articulated better.

## References

Barab, S., & Squire, K. (2004). 'Design-based research: Putting a stake in the ground', *Journal of the Learning Sciences*, 13(1), 1-14.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How People Learn: brain, mind, experience, and school*, Washington, DC: National Research Council.

Brown, A. L. (1992). 'Design experiments: Theoretical and methodological challenges in creating complex interventions', *Journal of the Learning Sciences*, 2, 141-178.

Cobb, P., & Bowers, P. (1999). 'Cognitive and situated learning perspectives in theory and practice', *Educational Researcher* 28(2), 4-15.

Collins, A., Joseph, D., & Bielaczyc, K (2004). 'Design research; Theoretical and methodological issues', *Journal of the Learning Sciences*, 13(1), 15-42.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. (2007). *Taking science to school: Learning and teaching science in grades K-8*, Washington, D.C.: National Academy Press.

Harrison, A., Klease, G., & Voss, L. (April, 2006). *Chemical literacy in grades 7-12: Teaching particle concepts for understanding*. A paper presented at the annual conference of the National Association for Research in Science Teaching, San Francisco, CA.

Krajcik, J.S., McNeill, K. L., & Reiser, B. J. (2007). *Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy*. Retrieved Nov. 17, 2007, from www.interscience.wiley.com.

Messick, S. (1994). 'The Interplay of Evidence and Consequences in the Validation of Performance Assessments', *Educational Researcher*, 23(2), 13-23.

Mislevy, R. J., & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology*, Menlo Park, CA: SRI International.

National Assessment Governing Board. (2006a). *Science assessment and item specifications for the 2009 National Assessment of Educational Progress* (Prepublication ed.), Online. Available HTTP: <http://www.nagb.org/ pubs/naep science specs 2009.doc> (accessed 11 July 2006).

National Assessment Governing Board. (2006b). *Science framework for the 2009 National Assessment of Educational Progress* (Prepublication ed.), Online. Available HTTP: <http://www.nagb.org/pubs/ naep science framework 2009.doc> (accessed 11 July 2006).

National Research Council (2006). *Systems for state science assessment*, Washington, DC: The National Academies Press.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*, Washington, DC: National Academies Press.

Scalise, K., & Gifford, B. R. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *Journal of Teaching, Learning and Assessment, 4(6)*.

Shin, N., Stevens, S. Y., & Krajcik, J. (2010). "Tracking student learning over time using Construct-Centered Design." In S. Routledge (Ed), *Using analytical frameworks for classroom research: collecting data and analysing narrative*. Taylor & Francis, London.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). 'Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic molecular theory', *Measurement: Interdisciplinary Research and Perspectives,* 4 (1), 1 – 98.

Stevens, S. Y., Delgado, C., & Krajcik, J. S. (in press). 'Developing a theoretical learning progression for atomic structure and inter-atomic interactions', *Journal of Research in Science Teaching*.

Stevens, S. Y., Sutherland, L. M. and Krajcik, J. S. (2009). 'The big ideas of nanoscale science and engineering'. Arlington, VA: NSTA Press.

Treagust, D. F., & Chandrasegaran, A. L. (2007). 'The Taiwan National Science Concept Learning Study in an International Perspective', *International Journal of Science Education*, 29 (4), 391 — 403

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum.

Wilson, M., & Berenthal, M. W. (2006). *Systems for state science assessment*. Washington DC: National Academies Press.

## Acknowledgments

This research is funded by a Developing an Empirically-Tested Learning Progression for the Transformation of Matter to Inform Curriculum, Instruction and Assessment Design grant, number 0822038, and a National Center for Learning and Teaching in Nanoscale Science and Engineering grant, number 0426328, from the National Science Foundation. Any opinions expressed in this work are those of the authors and do not necessarily represent those of the funding agency.

# Discourse as a lens for reframing consideration of learning progressions

Alicia C. Alonzo, Michigan State University, 330 Erickson Hall, East Lansing, MI 48823, alonzo@msu.edu

**Abstract:** Learning progressions – descriptions of increasingly sophisticated ways of thinking about a topic – have generated much interest in the science education community, as a means of bridging the chasm between cognitive science research and methods for supporting and assessing student learning. In order to fulfill this promise, learning progressions must be developed and refined based upon a rigorous process that explores the match between the learning progression and available evidence of student learning and is complicated by the nature of students' thinking. This paper explores the implications of reframing consideration of learning progressions using Gee's notion of discourse. This reframing is posited as a means of making sense of students' use and interpretation of words that have both everyday and scientific meanings and their apparent inconsistency in responding to assessment items. Data collected about students' understandings relative to two learning progressions is used to explore this approach.

Learning progressions – descriptions of increasingly sophisticated ways of thinking about a topic (National Research Council [NRC], 2007) – have generated much interest in the science education community. The National Science Foundation (NSF) has issued special calls for projects focused on learning progressions (NSF, 2005, 2006, 2009); the National Research Council has advocated for their use to inform science curricula (NRC, 2007) and assessments (NRC, 2006); and the National Assessment of Educational Progress (NAEP) science framework called for the inclusion of learning-progression-based items (National Assessment Governing Board, 2008). Learning progressions have been promoted as solutions to wide range of current educational issues, including a lack of curricular coherence (Gotwals & Songer, 2008), the developmental inappropriateness of current curricula (Stevens, Shin, Delgado, Krajcik, & Pellegrino, 2007), misalignment between instruction and assessment (Black & Wilson, 2007), and weaknesses in support for valued teaching practices (Alonzo, 2009; Furtak, 2009). Learning progressions are viewed as a means of bridging the chasm between cognitive science research and methods for supporting and assessing student learning.

In order to fulfill this promise, learning progressions must be developed based upon existing research and refined based upon a rigorous validation process. The central question guiding this work is the extent to which the learning progression (as a hypothesis about student learning) matches available evidence. This evidence necessarily relies upon methods for eliciting and interpreting student responses relative to the learning progression. Thus, the iterative process of developing and refining a learning progression must involve simultaneous consideration of both student learning and the means by which that learning is assessed. Both the learning progression and its associated assessment items are influenced by assumptions that are made about what is progressing and the goal of learning.

This paper explores an alternative framing of these assumptions, as a means of addressing two current challenges in work to describe and assess student learning through the use of learning progressions. The paper first provides an overview of these challenges. Second, the paper describes an alternative framing – a different way of considering what is progressing – drawing upon Gee's (1991) notion of discourse. This framing is then applied to consider evidence of student learning. The paper concludes with a discussion of implications for future work on learning progressions.

Throughout this paper, two learning progressions are used to illustrate these ideas. The first was designed to describe the learning that might take place during students' first formal study of *Force & Motion* (*FM*; Alonzo & Steedle, 2009), typically in a middle or high school physical science course. It focuses upon students' ability to make predictions about situations involving force and/or motion: predicting the force(s), given information about an object's motion, and vice versa. The second learning progression describes elementary school students' learning about *Plant Nutrition* (*PN*; Alonzo, Benus, Bennett, & Pinney, 2009). It is comprised of three "progress variables" (Wilson, 2005): *Food for Plants*, *Energy for Plants*, and *Plants as Producers*. The discussion in this paper is limited to the first two progress variables, which pertain to students' explanations surrounding the source of plants' food and energy, respectively.

## Two Challenges in Current Work on Learning Progressions

Work to develop learning progressions (and associated assessment items) must grapple with the nature of student thinking. Prior work (Alonzo & Steedle, 2009) explored two characteristics of students' thinking that pose challenges to efforts to describe and assess their learning. Each of these challenges will be briefly described here and illustrated with examples from the two learning progressions.

## Students' Definitions for Words with Scientific and Everyday Meanings

In many areas of the science curriculum, students encounter words that have both "scientific" and "everyday" meanings. For example, the word "force" has a variety of everyday meanings (Halloun & Hestenes, 1985) across a range of different contexts. A sports fan might consider ways in which this word is used as a verb – for example, forcing a double play [in baseball], a game four [in a playoff situation], or a turnover [in basketball or football] – and as a noun, referring to a player as "a force" [such as a defensive lineman in football]. These examples might lead one to conclude that force has a meaning like "making something happen" (or stopping something from happening), a much broader definition than that of a physicist. Ioannides & Vosniadou (2001) explored the ways in which children from 4 to 15 years of age used the word "force". They identified four well-defined, internally consistent meanings for this word and found that students' use of these definitions varies with age. Thus, it seems likely that students' ideas about the word "force" change along with their ideas about the content of the *FM* learning progression: the relationship between (the scientist's definition of) force and motion.

Revisions to the *FM* learning progression (Alonzo & Steedle, 2009) included consideration of the way in which students' definitions for the word "force" were treated. Students in the target population were expected to hold an "acquired force" (Ioannides & Vosniadou, 2001, p. 26) or "impetus" view of force (Buridan, as quoted in Halloun & Hestenes, 1985, p. 1057): the idea that moving objects have some internal property, imparted to them by a push or a pull. Typical physics textbooks (e.g., Hewitt, 1997) introduce a scientific definition for force and then assume students will use that definition to make sense of subsequent discussion of the relationship between force and motion. Similarly, initial versions of the learning progression first addressed students' definitions for the word force, such that a lower level (Level X on the left side of Figure 1) was reserved for students who hold an impetus view of force. Higher levels of the learning progression (Levels X+1 and X+2 on the left side of Figure 1) described increasingly sophisticated ideas about the relationship between force and motion, assuming that students at those higher levels hold a scientific view of "force".

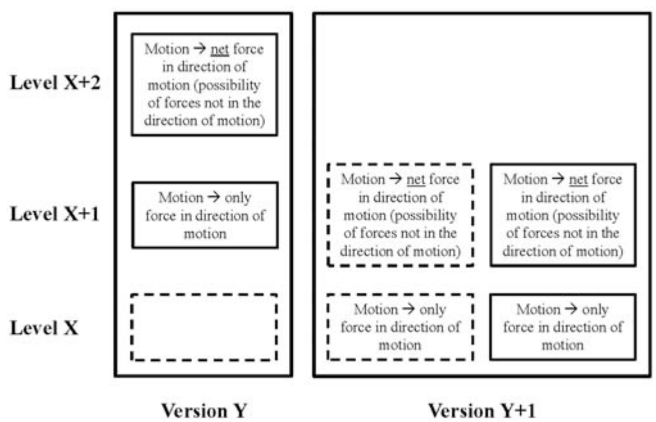



Figure 1. Revisions to the *FM* learning progression, reflecting students' definitions for the word "force". Dashed boxes represent an "impetus" view of force: solid boxes represent a scientific view of force.

Perhaps not surprisingly, interview data revealed that students with an impetus view of force could have different ideas about the relationship between force and motion (Alonzo & Steedle, 2009). For example, when asked to identify the forces acting upon a ball thrown up into the air, students might say that only the throwing force is acting on the ball or that both the throwing force *and* gravity are acting on the ball. The first response is consistent with the view that the only force acting on an object is one in the direction of its motion. The second response is slightly more sophisticated, recognizing the importance of *net* force and the possibility of forces acting on an object that are not in the direction of its motion. Both of these ideas about the relationship between force and motion could also be held by students with more scientific views of the word "force". Thus, as shown on the right side of Figure 1, the learning progression was revised to reflect the possibility of students with two different views of "force" making progress in thinking about the relationship between force and motion, such that students' ideas about "force" and their use and interpretation of this word were treated as two separately progressing entities.

Work on the second learning progression was similarly complicated by students' definitions for words such as "food" and "energy". A review of the literature indicated that students often use these words interchangeably (Leach, Driver, Scott, & Wood-Robinson, 1996) to describe things that are necessary for plants to be healthy. Students who do differentiate "food" and "energy" tend to equate food with things that are "eatable" (Wood-Robinson, 1991; Simpson & Arnold, 1982). Initial data collection surrounding this learning progression elicited evidence of students' definitions for these words, along with evidence of their thinking about plant nutrition (Alonzo et al., 2009). Once again, it appeared that students could make progress with respect to underlying ideas about *Food for Plants* and *Energy for Plants*, while holding a range of different definitions for words such as "food" and "energy". The work reported in this paper was motivated by an interest in this central problem: the way in which students' definitions for words (particularly those with both everyday and scientific meanings) impact their learning and expression of scientific ideas.

## Apparent Inconsistency of Students' Thinking

In physics, students' inconsistency in approaching problems addressing the same underlying principles is well-documented. Chi, Feltovich, and Glaser (1981) demonstrated that a key characteristic differentiating between novices and experts is the way in which they categorize physics problems, with experts viewing problems in terms of underlying principles and novices attending to surface features. Halloun and Hestenes (1985) and Finegold and Gorsky (1991) identified frameworks that college-level students used to reason about motion and force, respectively; however, students did not apply these frameworks consistently across different problem situations.

Reinforcing these findings from the literature, inconsistencies were observed in student responses elicited to explore the *FM* learning progression (Alonzo & Steedle, 2009). For example, a student might describe the forces acting on a stationary object (such as a book on a table), by explaining that "gravity is always acting." Yet this same student might respond that only a throwing force is acting on a ball tossed into the air, rejecting the notion that gravity is acting on the moving ball. Reliabilities for sets of items assessing understanding with respect to the learning progression were all moderate. In addition, a relatively low percentage (59%) of student responses were at the same level across pairs of items that asked essentially the same question in two different contexts. Similarly, in interviews designed to inform revision of the *PN* learning progression (Alonzo et al., 2009), inconsistencies were observed in the way in which students discussed the source of food for plants. At the beginning of the interviews, students were asked to describe the different parts of a plant and to explain what each part did "to help the plant live and grow." Many students began by explaining that the roots take in food for the plant. Yet later in the interview, these same students might express the idea that plants "make" or "produce" their own food.

Thus, listening to students discuss a set of force and motion problems or talk about "food for plants", it can be easy to conclude that students' thinking simply doesn't make sense. Although the work described in this paper was not undertaken to address this puzzle, as will be discussed below, attempts to make sense of students' use of language had the unexpected, but fruitful, side effect of also suggesting an alternative interpretation for the apparent inconsistency of students' thinking.

## Reframing Thinking about Learning Progressions in Terms of Discourse

Gee (1991) defines discourse as "a socially accepted association among ways of using language, of thinking, and of acting that can be used to identify oneself as a member of a socially meaningful group or 'social network'" (p. 3). Elizabeth Moje and colleagues (2004) have explored the implications of this perspective for literacy and content instruction in science classrooms; this paper explores implications for describing and assessing student learning via learning progressions.

Returning to a consideration of everyday meanings for the word force, it becomes evident that these definitions, while inconsistent with a scientist's definition, are completely consistent with one's everyday experiences with force and motion. In everyday experience, forces are required to keep objects moving. A force is, indeed, required to make something happen (in this case, the motion of an object). One might conceptualize the whole "package" of one's ideas, experiences, and ways of talking as comprising an "everyday discourse" with respect to force and motion. We know a lot about words such as "force", "food", and "energy" by virtue of the way in which we use them to talk about experiences in our everyday lives. Those definitions do not exist in isolation; rather, they are a part of how we "are in" and interact with the world and, thus, inform the way in which we think about related ideas (such as the relationship between force and motion or plant nutrition). While these ideas are perhaps not a central part of our identities, the particular experiences we have in the world – by virtue of being, for example, a sports fan – do influence our thinking about words that have both everyday and scientific meanings and, thus, are tied up with how we make sense of the world.

This "everyday discourse" can be contrasted with a "scientific discourse".[1] The latter includes the use of words consistent with their scientific definitions, as well as the application of a set of scientific ideas derived from those definitions to very specific sorts of situations. In many cases, this means accepting a narrower view of words and ideas than those in one's everyday experience. For example, a physicist would not say that forces make "something" happen. He/she would use force only to apply to case in which "something" refers to a change in motion (speed and/or direction). The scientific discourse also requires – to a certain extent, at least in physics – that one reside in a fairly idealized world: one in which friction and air resistance are usually ignored.

Taking a discourse perspective recasts what is progressing in a learning progression in terms of a transition between "everyday" and "scientific" discourses. In their current forms, the *FM* and *PN* learning progressions describe the evolution of students' thinking about a scientific topic. Students' definitions for relevant words, such as force, food, or energy, are treated as existing separately from their understanding of the scientific topic and as a challenge to accessing the real focus of the learning progression. Taking a discourse perspective puts students' definitions and ideas back together again. This reframing also permits a reinterpretation of students' seemingly inconsistent responses as struggles to reconcile their everyday and scientific discourses. Sometimes there may not be a conflict between these two discourses. However, where everyday and scientific discourses yield different ways of interpreting and reasoning about a given question, students must choose how they will consider the question being asked. Subtle aspects of the situation – such as the context in which questions are being asked, the particular problem context involved in a given question, and even particular wording of a question – might influence the extent to which students draw upon their everyday and scientific discourses in making sense of the situation.

## Reconsidering Evidence of Student Learning

In this section, data collected with respect to the *FM* and *PN* learning progressions are reconsidered in light of the discussion above, in order to answer the research question, "Does a discourse lens help to interpret students' responses with respect to the two learning progressions?" This is intended as a "proof of concept" exploration, as the data were not collected specifically to test the ideas above. Rather, the intent is to explore the fruitfulness of this reframing, in order to inform future data collection that might more directly evaluate these ideas.

### Evidence with Respect to the *FM* Learning Progression

The data considered in this section was collected as part of two separate studies, both conducted in the Midwestern United States. Study 1 involved students from a middle class rural high school who answered questions about force and motion directly following a unit on this topic in their physical science classes. Data included 58 students' responses to ordered multiple choice (OMC; Briggs, Alonzo, Schwab, & Wilson, 2006) items. Study 2 involved 1088 students from six high schools, representing a mix of rural and suburban communities. By virtue of their course enrollment, students were expected to have learned about force and motion concepts (either in their current science course or one taken in a previous year). Data included students' responses to a set of OMC force and motion items, plus one question that asked students to indicate whether they had learned about the content of those items before.

The source of students' reasoning about the OMC items was not evident in their responses. However, one might hypothesize that students who have access to the scientific discourse are more likely to reason using that discourse in situations further removed from their everyday experiences. The available data permits a small exploration of this hypothesis. In particular, the set of 16 items presented to students in both Study 1 and Study 2 include three pairs of items in which an almost identical question is asked in a familiar and an unfamiliar context. Questions A and B ask students to describe the motion of an object, first under the influence of a constant force and then once that force is no longer being applied, respectively. These questions are asked both in the relatively familiar context of pushing a car (albeit across a frictionless surface) and the less familiar situation of a rocket traveling in outer space. Question C asks students to describe the forces acting upon a stationary object, both a boulder resting upon the ground (familiar context) and a mysterious blob suspended in the air in a scientific laboratory (unfamiliar context). Analysis of responses to these six questions involved the 58 students from Study 1 and 798 students from Study 2 (those who indicated that they had learned about force and motion concepts before). Students who had not studied force and motion were excluded, as these students may not have access to the scientific discourse and, thus, would not be expected to exhibit variation in the discourse used to reason about the different problem contexts.

Table 1 shows the percentage of students in each study choosing the option corresponding to the highest learning progression level (as an indicator of the use of scientific discourse) for each item. Shaded cells indicate statistical significance ($p < .05$), using the McNemar test. For all comparisons, students selected the highest learning progression level at a similar or higher rate in the familiar context, as compared to the unfamiliar context. However, these rates were statistically significant for only two pairs of items (Question B in Study 1 and Question C in Study 2). Thus, across the two studies, students who had studied force and motion concepts seemed somewhat more likely to apply scientific discourse to analysis of unfamiliar contexts, lending a small degree of support to the proposed reframing.

Table 1: Percentage of students selecting the highest learning progression level for a given item.

| | Study 1 (N=58) | | Study 2 (N=798) | |
|---|---|---|---|---|
| | Familiar Context | Unfamiliar Context | Familiar Context | Unfamiliar Context |
| Question A | 17.2% | 22.4% | 16.5% | 17.3% |
| Question B | 31.0% | 53.4% | 40.4% | 40.2% |
| Question C | 43.1% | 53.4% | 37.8% | 60.0% |

## Evidence with Respect to the *PN* Learning Progression

In order to inform revisions of the *PN* learning progression, semi-structured interviews were conducted with 30 students (grades 1-6) who attended a suburban elementary school in the Midwestern US. Because of its proximity to a large research university and hospital, the school enrolled many children whose parents were associated with these institutions. The school was located on the edge of town and, thus, some of its students lived on surrounding farmland. A small houseplant was used as a prop during the interviews, and efforts were made to elicit students' ideas about words such as "food" and "energy", as well as their school-based and everyday experiences with plants. However, the richest interview data came from students who spontaneously explained their answers in terms of their own experiences and everyday discourses.

Interviews were labeled with the gender and grade level of the student, as well as the order in which they were conducted. For example, 04-M1 indicates the fourth interview, conducted with a male first grader. All interviews were videotaped, and the videotapes were transcribed prior to analysis. For the purposes of this paper, interview transcripts were coded to identify instances in which students explicitly provided evidence of their everyday or scientific discourses; this evidence included explicit use of the words "food" and "energy" as well as students' discussion of related ideas. Students' examples of how they used words such as "energy" in their everyday lives, descriptions of out-of-school experiences with plants, and explanations for personal experiences that informed their thinking about plants were all considered to be evidence of everyday discourse. Students' discussion of ideas explicitly identified as having been learned in school or from a science text were considered to represent their scientific discourse. The surrounding interview transcript was then examined for ways in which students' ideas about plants' sources of food and energy may have been influenced by these two types of discourse. Themes emerged with respect to each of the two progress variables and are illustrated with examples from the interview transcripts.

### Food for Plants

Consistent with findings from other research (e.g., Cañal, 1999), the students in this sample often expressed the view that plants manufacture their own food along with the view that plants take in food from the environment. Some students expressed each idea separately, at different points in the interview, while other students stated that plants get food both from sources. Although the latter position is represented in the current version of the *Food for Plants* progress variable, a student who exhibits the former type of response to the interview questions appears to be at one level of the progress variable at some times and at a completely different level at other times. Such a student poses challenges to current conceptualizations of learning progressions, which assume that a given student will hold a consistent set of ideas that can be used to diagnose his/her level. However, since each view about the source of food for plants is consistent with a different discourse, students' thinking may be reinterpreted as struggles to reconcile two different ways of thinking and talking about food for plants.

Although most students did not explicitly address the source of their ideas about plants manufacturing their own food, this is unlikely to be a part of most students' everyday experiences and, thus, may be interpreted as representing a scientific discourse. For example, a second grade boy's interview (20-M2) included the following statement about the role of the sun in plants' growth: "I think plants use the sun to make their food… I think in the photosynthesis they do that." Later in the interview, he explained that he had read about photosynthesis in a science book. Through their school science experiences, many students in the interview sample had developed a discourse surrounding seeds and their source of food. For example, a third grade girl (02-F3) explained that she had learned in school that "the seed doesn't need food 'cause they have already food… they have food in their shell."

Many of the same students also had everyday discourses surrounding food and, in particular, plants' source for food. Consistent with Roth's (1991) interpretation of student thinking, some students brought in ideas about human nutrition to explain their ideas about plant nutrition. For example, a third grade girl (17-F3) said, "We grow because we eat healthy food… Like, if we ate candy bars, like, every single day, like ten a day. We would get really, really sick. And we wouldn't, like, be very healthy. And we wouldn't grow as well." She went on to explain that it wouldn't be healthy for plants to take in "a kind of bad soil that's really sugary." Two students (04-M1 and 21-M5) discussed food for plants in terms of experiences with Venus flytraps, which "eat bugs" that they fed to the plants. The first grade boy (04-M1) had difficulty reasoning about food for regular plants, which he recognized did not eat bugs. Most students' out-of-school experiences with plants were situated in a discourse about plants needing things from people and those things constituting food for plants. For example, a third grade girl (02-F3) described in vivid detail her experiences planting trees on her family's farm over the weekend preceding the interview. She had provided the new trees with both "special food" and water (that "carries food in it"). For other students, including another third grade girl (17-F3), the water itself, as something that was required for life, was also food for the plant. She explained that this saying, "'Cause just like people, they need water to survive." For these students, ideas about human nutrition and human interactions with plants (part of their everyday discourse) fairly straightforwardly informed their ideas about plants.

However, the interviews also offered numerous examples of students who appeared to be using different discourses to make sense of the different questions being asked about food for plants. Many of these students appeared to be struggling to reconcile everyday and scientific discourses. Often, student's scientific discourses did not appear to be robust enough to support detailed reasoning about food for plants. Students seemed to have learned a phrase, such as "plants produce their own food" without details about this process. For example, consider the following segment of interview transcript involving a second grade girl (02-F3):

> 02-F3: [Plants] produce the food.
> Interviewer: Do they use any of the food that they produce?
> 02-F3: What do you mean "they produce"?

The idea of plants producing food seems to have limited meaning to the girl. When her ideas about plants' source of food was probed further in the interview, she seems to have combined ideas about plants taking in food and water (provided by people such as herself) to explain that "The water actually goes to the leaves and the leaves can turn it into food. The leaves could have to make the food out of the soil. The water too." Thus, like many students in the sample, she appears to be reconciling her ideas about plants taking in food and water from the environment (to serve as food) with the idea of plants making their own food.

## Energy for Plants

Perhaps not surprisingly, students in the interview sample seemed to have limited access to a scientific discourse surrounding energy. However, they provided many examples of their everyday discourse about energy, allowing examination of the way in which this discourse informed their thinking about energy for plants. Students expressed three main ideas about energy – associating energy with movement, electricity, and the sun – all of which were contained in a segment of an interview with a third grade girl (17-F3):

> Well, energy can come from the sun. Energy is, like, a thing that makes you move. Like, if you're really hyper you can have a lot of energy so you can, like, run around a lot. And energy is, like, batteries have energy in them. And electronics. And, energy from the sun also, like, it has vitamins in it. Kind of. So that, like, when it comes down and hits stuff, it makes it healthy.

Thus, consistent with other research (e.g., Nicholls & Ogborn, 1993), students' everyday discourses about energy often involved movement. These students then tried to make sense of energy for plants in terms of this way of thinking about energy. One second grade girl (14-F2) used her ideas about energy ("Like if you run then you kinda get your energy") to conclude that plants would not have energy "unless someone moved it… If someone picked it up and ran with it, maybe it would have energy." A third grade girl (02-F3) explained that plants have energy "not in order to, from here, this plant to move from here to here. Not that type of energy." She later explained that plants needed energy so "the roots can move around to get the food." Thus, she associated plant's energy with motion of its roots. Other students invoked the movement of substances such as food or water to explain how plants use energy. For example, a fourth grade boy (01-M4) explained that plants might use energy "to get the food to the, everywhere it needs to go… Like the leaves."

Students also drew upon their own experiences in thinking about where plants might get energy. For example, the third grade student above (17-F3), like other students in the sample, believed that the sun was a source of vitamins for both people and plants. She also explained that people get energy from healthy food:

> You get your energy from food. Like, carrots, and celery, and, um sugar gives you energy, but it's a bad kind of, well, it's not really bad, but it just, like, doesn't last very long. So, it's not, like, true energy. But, food that has, like, vitamins in it, I think it has, like, the vitamins and the calcium combined, that gives you energy. So, that, like, if you eat a bunch of carrots, and celery, and vegetables, and fruits, that you'll be really energetic because then you have a lot of… you have a lot of nutrients in you. And so then you're like, really healthy."

She went on to explain that plants needed healthy food and vitamins for their energy as well. A second grade girl (05-F2) explained her knowledge of energy in terms of her mom's position as a track coach:

> [Energy] means exercise and fit, so you can stay in shape and stuff. 'Cause I know all this stuff because… my mom's a track coach so she teaches about 85 people in this team. And she has to tell them about keep fit and stay in shape and stuff.

Later in the interview, she explained that she knew plants get energy from water (as well as sunlight) "'cause we get energy from water, 'cause yesterday I ran a mile. And I needed a lot of water."

The students in the interview sample had not had many formal learning opportunities surrounding the concept of "energy", but they drew upon rich everyday discourses to make sense of energy for plants. This led them to make connections between plants and humans, some of which are consistent with a scientific discourse about energy for plants and some of which are not. Thus, we might predict that their subsequent learning about plants will involve some struggle to identify aspects of their everyday discourse that are and are not consistent with scientific discourse.

## Conclusions & Implications

The reframing presented in this paper takes as an underlying assumption that science learning involves students' reconciling their everyday and scientific discourses and that this process will occur whether or not it is acknowledged in efforts to describe, support, and assess student learning. While students' prior conceptions are commonly acknowledged, this perspective takes a broader view of students' ideas, explicitly including ways in which students draw upon both language and experiences to make sense of phenomena. The preceding sections of this paper explored whether this assumption was reflected in existing data surrounding students' understanding of force and motion and food and energy for plants. Although the data was not collected to explicitly address students' everyday discourses, the available evidence seems to support the interpretation that students draw upon this discourse in making sense of questions about scientific topics, often in conjunction with the scientific discourse. This interpretation helps to make sense of both students' use of language and their apparent inconsistency in responding to assessment items and interview questions.

This perspective has several implications for the design of learning progressions. First, treating students' ideas about and language surrounding scientific topics as both falling under the "umbrella" of discourse helps to address the perhaps false dichotomy between ideas and language. This helps to resolve the problem of how to "treat" students' definitions for words as part of a learning progression. Second, this approach means reframing what is progressing in a learning progression from ideas about a scientific topic to changes in the extent to which students draw upon scientific discourse in reasoning about scientific questions. It is important to recognize that this process is not one of replacement, in which students make progress by gradually discarding their everyday discourses in favor of a scientific one. Instead, one might view this as a process by which students gradually understand their everyday discourses in light of a scientific one: narrowing definitions and ideas where necessary and adding additional interpretations to explain existing experiences. As students make progress with respect to the learning progression, they may draw upon everyday and scientific discourses to varying degrees; part of what is progressing is students' reliance upon scientific discourse in appropriate situations. Thus, in this view, one does not expect students to consistently rely upon scientific discourse to answer all questions and – to both describe and support student learning – a challenge becomes figuring out the circumstances under which students might be more likely to draw upon everyday discourse and to help them to utilize scientific ways of thinking in appropriate circumstances. The reframing proposed in this paper has the potential to help learning progressions to more accurately reflect the nature of students' thinking and, thus, to be more useful tools for describing, assessing, and supporting student learning.

## Endnotes

(1) This discussion is not meant to imply that everyday and scientific discourses are fully dichotomous. Scientific discourse emerges from and attempts to explain everyday experiences, and, indeed, one would expect a scientist (or expert student) to have access to both types of discourse and to flexibly draw upon some mixture of the two, depending upon the context. One might view these two types of discourse as two ends of a continuum (where the continuum is shorter or longer, depending upon the degree of similarity between everyday and scientific discourses in a particular area).

## References


Alonzo, A. C. (2009, April). *Design criteria for learning progressions to support teachers' formative assessment practices.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education, 93*, 389-421.

Alonzo, A. C., Benus, M., Bennett, W., & Pinney, B. (2009, September). *A learning progression for elementary school students' understanding of plant nutrition.* Paper presented at the biennial meeting of the European Science Education Research Association, Istanbul, Turkey.

Black, P., & Wilson, M. (2007, April). *The idea of a learning progression as a core for both instruction and assessment*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment, 11*, 33-63.

Cañal, P. (1999). Photosynthesis and "inverse respiration" in plants: an inevitable misconception? *International Journal of Science Education 21*, 363-371.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Characterization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Finegold, M., & Gorksy, P. (1991). Students' concepts of force as applied to related physical systems: A search for consistency. *International Journal of Science Education, 13,* 97-113.

Furtak, E. M. (2009, April). *Learning progressions to support teacher development.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Gee, J. P. (1991). What is literacy? In C. Mitchell & K. Weiler (Eds.), *Rewriting literacy: Culture and the discourse of the other.* New York: Bergin & Garvey.

Gotwals, A. W., & Songer, N. B. (2008, March). *Scenario-based assessment: Gathering evidence of student reasoning in biodiversity.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Halloun, I. A., & Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics, 53*, 1056-1065.

Hewitt, P. G. (1997). *Conceptual physics* (3rd ed.). Menlo Park, CA: Addison-Wesley Publishing Company.

Ioannides, C., & Vosniadou, S. (2001). The changing meanings of force: From coherence to fragmentation. Cognitive Science Quarterly, 2(1), 5 – 62. Retrieved from http://www.cs.phs.uoa.gr/el/staff/vosniadou/force.pdf

Leach, J., Scott, P., Driver, R., & Wood-Robinson, C. (1996). Children's ideas about ecology 2: ideas found in children aged 5-16 about the cycling of matter. *International Journal of Science Education, 18*, 19-34.

Moje, E. B., Ciechanowski, K. M., Kramer, K., Ellis, L.,Carrillo, R., & Collazo, T. (2004). Working toward third space in content area literacy: An examination of everyday funds of knowledge and Discourse. *Reading Research Quarterly, 39*(1), 38-70.

National Assessment Governing Board. (2008). *Science framework for the 2009 National Assessment of Educational Progress.* Retrieved from http://www.nagb.org/publications/frameworks/ science-09.pdf

National Research Council. (2006). *Systems for state science assessment*. Washington, DC: The National Academies Press.

National Research Council (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: The National Academies Press.

National Science Foundation. (2005). *Instructional Materials Development program solicitation* (NSF 05-612). Arlington, VA: Author. Retrieved from http://www.nsf.gov/pubs/2005/nsf05612/ nsf05612.htm

National Science Foundation. (2006). *Discovery Research K-12 program solicitation* (NSF 06-593. Arlington, VA: Author. Retrieved from http://www.nsf.gov/pubs/2006/nsf06593/nsf06593.pdf

National Science Foundation. (2009). *Research and Evaluation on Education in Science and Engineering program solicitation* (NSF 09-601). Arlington, VA: Author. Retrieved from http://www.nsf.gov/pubs/2009/nsf09601/nsf09601.txt

Nicholls, G., & Ogborn, J. (1993). Dimensions of children’s conceptions of energy. *International Journal of Science Education, 15*, 73-81.

Roth, K. J. (1991). Learning to be comfortable in the neighborhood of science: An analysis of three approaches to elementary science teaching. In W. Saul & S. A. Jagusch (Eds.), *Vital connections; Children, science, and books* (pp. 143-161). Washington, DC: Library of Congress.

Simpson, M. & Arnold, B. (1982). Availability of prerequisite concepts for learning biology at certificate level. *Journal of biological education, 16*(2), 65-72.

Stevens, S.Y., Shin, N., Delgado, C., Krajcik, J., & Pellegrino, J. (2007, April). *Developing a learning progression for the nature of matter as it relates to nanoscience.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ” Lawrence Erlbaum Associates.

Wood-Robinson, C. (1991). Young people’s ideas about plants. *Studies in Science Education, 19*, 119–135.

## Acknowledgments

Initial development of the *FM* learning progression was conducted at the University of California, Berkeley with funding from the National Science Foundation (NSF) (#REC-0087848). Support for subsequent development and research with respect to this learning progression was provided by the NSF through a Center for Learning and Teaching grant to CAESL (Center for Assessment and Evaluation of Student Learning), the University of Iowa College of Education Research Fund, and the Iowa Mathematics & Science Education Partnership (IMSEP). Work on the *PN* learning progression was funded by IMSEP and the University of Iowa Department of Teaching & Learning. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author and do not necessarily represent the views of the NSF or other funding agencies.

# Examining the Role of Verbal Interaction in Team Success on a Design Challenge

Xornam S. Apedoe[1], Kristina V. Mattis[1], Bianca Rowden-Quince[1] & Christian D. Schunn[2]
[1]Department of Learning & Instruction, University of San Francisco, San Francisco, CA, 94117
[2]Learning Research & Development Center, University of Pittsburgh, Pittsburgh, PA 15260
xapedoe@usfca.edu; kvmattis@usfca.edu; bcrowden@usfca.edu; schunn@pitt.edu

**Abstract:** Collaboration is an essential component of many project-based learning environments; however, verbal interactions within groups can be ineffective. Extending previous research, this study investigated verbal interactions within teams of high school students as they solved a prototypical design-task called the Earthquake task. Videotaped performance assessments of high and low performing teams were analyzed in depth. Teams were categorized as high and low performing using two different methods: (1) highest successful structure built and, (2) percentage of successful designs. Results suggest that teams engaged in three different patterns of verbal interactions: *Accept-Build*, *Accept-Build-Discuss-Modify*, and *Discussion*. High performing teams most often engaged in, and were most successful when, their verbal interactions followed the *Accept-Build-Discus-Modify* pathway. The instructional implications of these results for promoting productive group collaboration and future research directions are discussed.

## Background

With the renewed interest in instructional approaches emphasizing the connection between knowledge and the context in which it is to be used, there has been resurgence in the use of project-based learning in K-12 science classrooms. The enthusiasm and belief in the efficacy for project-based learning has waxed and waned over the years, resulting in only a handful of teachers adopting the approach consistently (Barron et al., 1998). There is strong evidence that problem-and project-based learning can be successful (e.g. Cognition and Technology Group at Vanderbilt [CTGV]. 1994, 1997; Collins, Hawkins, & Carver, 1991; Schauble, Glaser, Duschl, Schulze, & John, 1995). Yet, there are several factors that are important for the design of problem-or project-based learning environments that can influence their impact on student learning outcomes.

Collaboration is an essential component of many project-based learning environments and can greatly impact student learning. It provides opportunities for students to extend their thinking, share ideas, and draw on the expertise of others (Krajcik et al., 1994). Yet, we have much to learn about how to best foster collaboration for learning (Barron, 2003). Simply having students work together to complete prescribed procedures for an investigation or task does not constitute collaboration (Marx, Blumenfeld, Krajcik, & Soloway, 1997). Rather, collaboration involves the construction of shared meanings (Webb & Palincsar, 1996) through the exchange of ideas.

Unfortunately, collaboration that is productive and leads to the development of shared understanding is not always guaranteed (Barron, 2003). Not all verbal exchanges within collaborative groups lead to the solving of problems or the construction of shared meanings. Rather, productive verbal interactions within a group require that: (a) ideas are presented clearly and explicitly so that they may become shared, (b) group members elaborate on their own and others' ideas, and (c) reasoning and evaluation of the ideas is done collaboratively (Hoek & Seegers, 2005; Mercer, 1996).

The nature and structure of the task can also shape the collaborative process (Hoek & Seegers, 2005). Tasks that are likely to benefit most from collaboration are those that require group work and deal with ill-structured problems (Cohen, 1994). Design is a practice that deals with ill-structured problems that have no clearly predefined path leading from the problem specification to the final design solution (Fortus, Krajcik, Dershimer, Marx & Mamlok-Naaman, 2005). For tasks, such as design tasks, productivity will depend on the interaction that happens within the group (Cohen, 1994).

In order to better foster productive collaboration within groups, we need to understand the characteristics of interactions that lead to productive group collaborations (Barron, 2003). Much like Barron (2003), this research is focused on understanding how verbal exchanges between collaborators influences the collective achievements of the group. While Barron focused on the interaction between collaborators while solving mathematical problems, this paper reports on the verbal interactions that occur within teams of students as they work to solve a design challenge. Design tasks provide an interesting context in which to study group collaborations because unlike math problems, which are well-structured, design problems are characteristically ill-structured, with ill-defined goals, states or operators (Goel & Pirolli, 1992). Thus, a goal of this research is to understand the verbal interactions that occur when students collaborate to solve a design challenge. Studies such as this are necessary if we are to begin to help practitioners and researchers create learning environments that can take full advantage of collaborative learning in project-based or design-based learning environments.

## The Study

This study is situated within the context of a larger study that examined student learning in an 8-week high school design-based learning curriculum unit. Design-based learning is a particular form of project-based learning. In design-based learning, the activity that is meant to promote learning is a design-project; students are required to use and extend their knowledge of science and math to develop a technological solution to a problem using available resources.

The research in this study focused on understanding how verbal interaction within teams affects performance in a design task. The analysis was motivated by prior research done within this context that examined the strategies students use to solve a prototypical design task, the Earthquake task (see Apedoe & Schunn, 2009). Results suggested that students use both science reasoning strategies (e.g. control of variables) and design-focused strategies (e.g. adaptive growth). However, the strategies commonly associated with success in science (e.g. control of variables) did not necessarily lead to success in design. In addition, while both science reasoning strategies and design-focused strategies led to content learning, the content learned was different. In this study, we extend the analyses to investigate the influence of verbal interactions on team performance on the Earthquake task.

The research questions of primary interest are:

1. What verbal interaction patterns are evident as teams collaborate to complete the Earthquake task?
2. What verbal interaction patterns distinguish high and low performing teams, as defined by the highest successful structure built?
3. What verbal interaction patterns distinguish high and low performing teams, as defined by the percentage of successful design trials?

### Research Design and Methods

#### Research Design

This study utilized a high-low extreme case design to analyze student team performances on a design task called the Earthquake task, which was adapted from Azmitia and Crowley (2001). Extreme-groups comparisons provide greater statistical power without bias, and are particularly useful for labor-intensive studies such as those involving video-data coding. Teams of students completed the task, and the most successful and least successful performances were analyzed in detail. There were two criteria used to categorize teams as high or low performing: height of structure, and percentage of successful designs trials regardless of height. Teams that were considered high performing based on the highest successful structure built may have taken more risks with their designs, whereas teams that had a high percentage of successful designs, but not necessarily tall structures, may have been more conservative in their approach. Thus it was of interest to see if the verbal interactions within teams would differ based on these two different categorizations.

#### The Earthquake Task

Students were given 54 wooden blocks to build as high a structure as possible that would withstand (i.e. no blocks fall from the structure) a 20-second simulated earthquake (see Figure 1). Students were allowed to build and test as many structures as they wanted and were given approximately 20 minutes to complete the task.

The Earthquake design task was chosen based on two important criteria. First, the Earthquake task has the three elements that Cross (1994) describes as being common to all design problems: (a) a specified goal, (b) constraints within which the goal must be achieved, and (c) criteria for recognition of a successful solution. Akin to the design problems described by Goel & Pirolli (1992), the Earthquake task is ill-structured and requires creativity to move from the start state to the goal state. Secondly, the task did not require students to have any discipline-specific knowledge about the domain beyond elements of everyday experience.

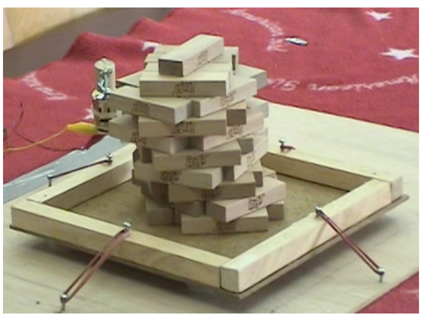


Figure 1. Sample student-designed earthquake-proof structure on earthquake simulator

## Context & Participants

This research was conducted as part of a larger study that examines student learning in high school design-based learning units. As part of an effort to reform teaching and learning in high school science classrooms, teachers from a large urban area in the northeastern United States were invited to implement a design-based science learning unit in their classrooms. Teachers held design competitions within their classrooms to identify the top student team design projects, as part of the unit implementation. These top student teams were then invited to participate in a *Regional Design Competition* (See Reynolds, Mehalik, Lovell, & Schunn, 2009 for more details). During the *competition event*, while teams waited for their projects to be judged, they were invited to participate in this study, and complete the Earthquake task. A total of 191 students (42% female, 58% male), drawn from urban high schools with a moderate proportion of traditionally underserved students, participated in the study. Participants included 9th, 10th, 11th and 12th grade high school students enrolled in a science course (e.g. spectrum science, biology, chemistry, or physics) that implemented a design-based learning unit. Participants worked in pre-existing groups of same or mixed gender teams of 2-5 students.

## Data Collection & Sources

The primary method of data collection for this study was performance-assessment interviews. Data were collected during the *Regional Design Competition* in both 2007 and 2008, with 29 teams of students participating in the 2007 session and 30 teams participating in the 2008 session. Student teams were videotaped while completing the Earthquake task. During the task, students were asked to explain their designs (e.g. Why did you build your structure this way?) to help clarify the design choices and problem-solving strategies they used. The data set analyzed for this paper included videos of 16 teams of students that represented both the highest and lowest performing teams based on the highest successful structure built. This sub-set was drawn from a larger dataset of 59 teams of students.

## Data Analysis

Video-data was analyzed in multiple passes by the researcher (Miles & Huberman, 1994). A coding scheme was developed to establish verbal interaction patterns within teams as they completed the design task. A set of 17 videos, that were not included in the final analysis, were viewed and coded. The final coding scheme developed showed that with the initial presentation of a design idea, the verbal interaction patterns within a team typically followed one of four pathways: *Accept Build*, *Accept Build-Discuss-Modify*, *Discussion*, or *Reject* (see Figure 2).

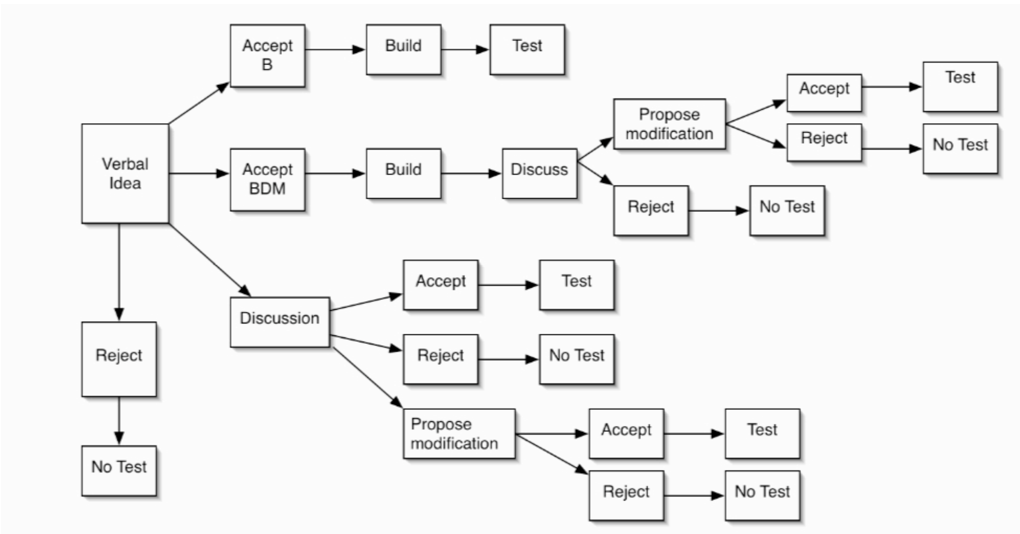



Figure 2. Possible Verbal Interaction Pathways

The first pathway, *Reject* describes the situation wherein a design idea is presented by a team member, but is immediately rejected by the rest of the team. The second pathway, *Accept Build (B),* describes instances wherein the initial design idea is immediately accepted and the team builds a structure.

The third pathway, *Accept Build-Discuss-Modify (BDM)*, describes instances wherein teams immediately begin building a structure. However, during the building of the structure, discussion ensues, often regarding the worthiness of the current design. At this point, either the current design idea is *rejected*, or a *modification* is proposed. This *proposed modification* is either accepted or rejected until an acceptable modification is proposed, and building resumes or a structure is tested. If an acceptable modification is not found, building stops, and the structure is never tested.

The fourth pathway, *Discussion*, begins with a design idea that is discussed by the team prior to physically building the structure. There are three sub-paths in the Discussion pathway: Discussion-Accept, Discussion-Reject, and Discussion-Proposed Modification. Discussion-Accept describes instances when a team member provides a design idea for the structure that is automatically accepted without any further discussion or modification prior to building and testing the structure. Discussion-Reject refers to the dismissal of a design idea after a discussion. In these instances, no structure is built until a new design idea is proposed and accepted. Discussion-Proposed Modification describes instances where a modification to the current design idea is suggested. These proposed modifications are either accepted or rejected. If accepted, the accepted proposed modifications are eventually tested after (n) accepted proposed modifications are made. If the proposed modification is rejected, then the team would continue to build the structure without any alterations, or until a new modification was proposed and accepted. All verbal interaction pathways end when the structure is finally tested to determine whether it could withstand the 20-second simulated earthquake.

Using the developed coding scheme, two coders independently coded the 16 teams' performances. Overall agreement between the two coders was 95.5%. In instances where there was a disagreement in coding, the discrepancy was discussed and resolved.

## Findings

### Verbal Interactions Used

All verbal interactions within teams were codable as one of the aforementioned pathways. Other logical possibilities, a *Reject* pathway and two sub-paths, the Discussion-Accept and Discussion-Reject, were never used by teams. Teams used the remaining verbal interaction pathways with varying frequency. The mean proportion of trials in which teams' verbal interactions followed the *Accept BDM* pathway was much greater than those that followed the *Accept B* or *Discussion* pathways. The average number of modifications made per trial during the *Accept BDM* pathway was 6.5, while during the *Discussion* pathway there was an average of 9 modifications per trial. Table 1 provides a summary of the overall trends in verbal interaction pathway use. Figure 3 shows the verbal interaction pathways used by teams.

Table 1: Summary of verbal interaction pathway use.

| Pathways | Mean Proportion of Use (all trials) | Mean Proportion of Successful Designs when Used (all trials) | Mean Proportion of Successful Designs (pathway specific trials) |
|---|---|---|---|
| Accept B | .23 | .18 | .53 |
| Accept BDM | .63 | .21 | .37 |
| Discussion | .14 | .40 | .50 |

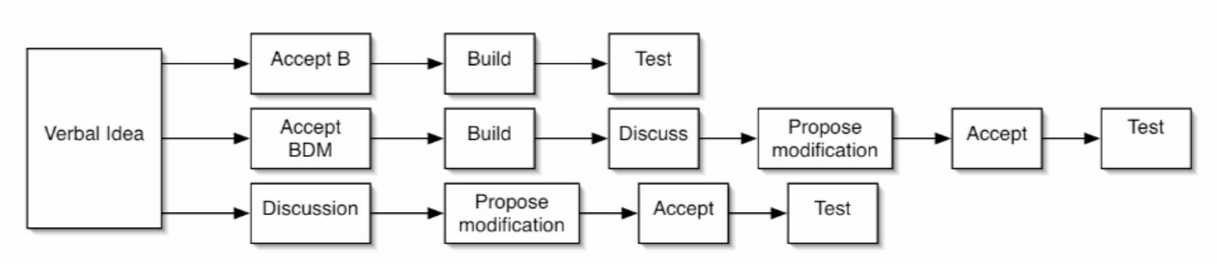



Figure 3. Observed Verbal Pathways Actually Used by Teams

### Distinguishing High & Low Performing Teams

#### Highest Successful Structure Built

Teams were categorized as high performing or low performing based on the highest successful structure they were able to build. Using a median split teams were divided into high and low groups, with the high groups' highest successful structures ranging between 13-45 stories tall, while the low group structures ranged from 0-10 stories tall.

High performing teams' verbal interactions followed the *Accept B* pathway significantly more frequently than did the low performing teams, $t$ (15) = 1.94, $p$ = 0.04. However, there was no significant difference in how frequently high and low performing teams' verbal interactions followed the *Accept BDM* pathway, $t$ (14) = -1.04, $p$ = 0.16 or the *Discussion* pathway, $t$ (8) = -0.44, $p$ = 0.33. Table 2 provides a summary of the verbal interaction pathways used by high and low performing teams.

With regard to design success, high performing teams had marginally significantly more successful designs across all trials when using the *Accept B* pathway, *t* (14) = 1.43, *p* = 0.09, as well as within trials that used the *Accept B* pathway, *t* (10) = 1.46, *p* = 0.09. High performing teams also had marginally significantly more successful designs when their verbal interactions followed the *Accept BDM* pathway over all trials, *t* (14) = 1.50, *p* = 0.08. When the analysis was narrowed to focus on only successful designs within trials that had verbal interactions following the *Accept BDM* pathway, high performing teams performed significantly better, *t* (14) = 1.90, *p* = 0.04. However, there were no significant differences in the amount of successful designs over all trials, *t* (14) = 0.83, *p* = 0.21, or within pathway only trials when verbal interactions followed the *Discussion* pathway *t* (14) = -0.61, *p* = 0.28. Table 3 provides a summary of the design success and its relation to verbal pathway use.

Finally, as part of the *Accept BDM* and *Discussion* pathways, the number of modifications proposed by team members was coded (see Table 3). There was no significant difference between high and low performing teams in the number of modifications made per trial when their verbal interactions followed the *Accept BDM* pathway, *t* (14) = 0.82, *p* = 0.21. There were also no significant differences in the number of modifications made per trial when their verbal interactions followed the *Discussion* pathway, *t* (12) = 0.59, *p* = 0.28.

Table 2: Verbal interaction pathway use by high/low teams as defined by highestsuccessful structure.

| | High Performing Teams (*M* proportion of trials) | Low Performing Teams (*M* proportion of trials) | Effect size (*d*) |
|---|---|---|---|
| Accept B* | .35 | .10 | 0.89 |
| Accept BDM | .55 | .72 | -0.51 |
| Discussion | .11 | .17 | -0.23 |

*Indicates statistically significant difference (p<0.05) between high and low performing teams

Table 3: Design success by verbal pathways for high/low teams as defined by highest successful structure.

| | Accept B | Accept BDM | Discussion |
|---|---|---|---|
| Successful Designs (all trials) | $M_{high}$ = .23▾<br>$M_{Low}$ = .06▾<br>*d* = 0.73 | $M_{high}$ = .52▾<br>$M_{Low}$ = .27▾<br>*d* = 0.75 | $M_{high}$ = .32<br>$M_{Low}$ = .17<br>*d* = 0.42 |
| Successful Designs (pathway specific trials) | $M_{high}$ = .31▾<br>$M_{Low}$ = .07▾<br>*d* = 0.73 | $M_{high}$ = .61*<br>$M_{Low}$ = .20*<br>*d* = 0.95 | $M_{high}$ = .13<br>$M_{Low}$ = .25<br>*d* = -0.29 |
| Modifications (per trial) | *N/A* | $M_{high}$ = 31<br>$M_{Low}$ = 22<br>*d* = 0.42 | $M_{high}$ = 6<br>$M_{Low}$ = 4<br>*d* = 0.23 |

* Indicates statistically significant difference (p<0.05) between high and low performing teams
▾ Indicates marginally statistically significant difference (p<0.10) between high and low performing teams

## Percentage of Successful Designs

Although success on the Earthquake task was defined as building as tall a structure as possible that would withstand a 20-second simulated earthquake, it was of interest to know whether teams who may not have built the tallest structures, but still had a high proportion of successful trials were engaged in different types of verbal interactions. Teams were once again evenly split into high and low performing groups, this time based on the percentage of successful designs they built and tested. The high performing group consisted of teams with a percentage of successful trials within the range of 38%-100%. The low group was comprised of teams with a percentage of successful trials within the range of 0%-33%. This categorization scheme resulted in a number of teams changing high/low status. Of the 16 total teams, 6 teams (38%) changed high/low status.

There were no significant differences between high and low performing teams in the frequency of use of the *Accept B, t* (14) = 0.19, *p* = 0.43, the *Accept BDM, t* (14) = -0.63, *p* = 0.27, or the *Discussion* pathways, *t* (10) = 0.61, *p* = 0.28. Table 4 provides a summary of the verbal interaction pathways used by high and low performing teams.

With regard to design success, there were no significant differences between high or low performing teams in the percent of successful designs over all trials when their verbal interactions followed the *Accept B* pathway, *t* (9) = 1.13, *p* = 0.14. However, there was a marginally significant difference in the proportion of successful trials within *Accept B* pathway only trials, *t* (8) = 1.44, *p* = 0.09. There was also a significant difference in the percent of successful design trials when the verbal interactions followed the *Accept BDM* pathway over all trials, *t* (10) = 3.62, *p* = 0.00. When the analysis was narrowed to focus only on the proportion of successful designs within trials that had verbal interactions that followed the *Accept BDM* pathway, high performing teams did significantly better, *t* (14) = 2.20, *p* = 0.02. For verbal interactions that followed the *Discussion* pathway, there was a marginally significant difference in the percent of successful designs over all trials, *t* (9) = 1.72, *p* = 0.06, however, there were no significant differences with the success of trials within *Discussion* pathway only trials, *t* (14) = 0.61, *p* = 0.28. Table 5 provides a summary of the design success and its relation to verbal pathway use.

Finally, there were no significant differences between high and low performing teams in the number of modifications made per trial when the verbal interactions followed the *Accept BDM* pathway, *t* (10) = 0.34, *p* = 0.37, or the *Discussion* pathway, *t* (10) = 0.61, *p* = 0.28, (see Table 5).

Table 4: Verbal interaction pathway use by high and low teams as defined by % of successful designs.

| | High Performing Teams (*M* proportion of trials) | Low Performing Teams (*M* proportion of trials) | Effect size (*d*) |
|---|---|---|---|
| Accept B | .24 | .22 | 0.07 |
| Accept BDM | .58 | .69 | -0.32 |
| Discussion | .18 | .10 | 0.30 |

Table 5: Design success by verbal pathways for high and low teams as defined by % of successful designs.

| | Accept B | Accept BDM | Discussion |
|---|---|---|---|
| Successful Designs (all trials) | $M_{high}$ = .21<br>$M_{Low}$ = .07<br>*d* =0 .58 | $M_{high}$ = .63<br>$M_{Low}$ = .16*<br>*d* = 1.81 | $M_{high}$ = .39˅<br>$M_{Low}$ = .10˅<br>*d* = 0.86 |
| Successful Designs (pathway specific trials) | $M_{high}$ = .31˅<br>$M_{Low}$ = .07˅<br>*d* = 0.74 | $M_{high}$ = .63*<br>$M_{Low}$ = .18*<br>*d* =1.12 | $M_{high}$ = .25<br>$M_{Low}$ = .13<br>*d* = 0.29 |
| Modifications (per trial) | *N/A* | $M_{high}$ = 28<br>$M_{Low}$ = 25<br>*d* = 0.14 | $M_{high}$ = 7<br>$M_{Low}$ = 3<br>*d* = 0.49 |

* Indicates statistically significant difference (p<0.05) between high and low performing teams
˅ Indicates marginally statistically significant difference (p<0.10) between high and low performing teams

## Discussion

The verbal interactions within both high and low performing teams followed one of three pathways, *Accept B*, *Accept BDM*, or *Discussion*. Overall, teams' verbal interactions followed the *Accept BDM* pathway. That is, once an initial design idea was posed, teams tended to begin building a structure that encapsulated the initial idea, and then as they were building, engaged in verbal exchanges wherein proposals for ways to modify the structure were made. Eventually, after a number proposed modifications were accepted or rejected, teams arrived at a structure that they deemed worthy of 'testing' to determine its ability to withstand the 20-second simulated earthquake. Overall, there was no statistically significant correlation between the verbal pathways used and the percentage of successful designs teams built.

However, when team performance was analyzed by looking at high performing and low performing teams, there appears to be a relationship between the verbal interaction pathway followed and how successfully teams were able to complete the design task. Teams that were considered high performing based on the highest successful structures they built engaged in verbal interaction that followed the *Accept B* pathway more often than did the low performing teams. In addition, the high performing teams had marginally significantly more successful designs when their verbal interactions followed the *Accept B* pathway than did the low performing teams. This result suggests that the verbal interaction pattern of: (1) proposing a design idea, (2) accepting the design idea, and (3) building the design idea without further discussion or modification of the idea prior to testing, can result in productive collaboration and success on the Earthquake task. The fact that the teams did not further discuss or modify the original design idea is the distinguishing difference between this pathway and the *Accept BDM* pathway during which modifications were made.

At first glance, this result is counter-intuitive because it would suggest that accepting and implementing ideas without discussion is a good strategy. However, we must be cautious when interpreting this result, as it is possible that high performing teams began with "better" ideas, and thus discussion or modification of those ideas was not necessary. In addition, teams that were considered high performing based on the highest successful structures they built, engaged in verbal interactions that did not follow the *Accept B* pathway the majority of the time. Rather, high performing teams most frequently engaged in verbal interactions that followed the *Accept BDM* pathway. Interestingly enough, high performing teams also experienced great success when their verbal interactions followed the *Accept BDM* pathway. High performing teams had significantly more successful designs when their verbal interactions followed the *Accept BDM* pathway. The primary difference between the *Accept B* and *Accept BDM* pathway is that within the *Accept BDM* pathway, teams engage in a cycle of proposing modifications to the initial design idea – and either accepting or rejecting those proposed modifications – before completing building of the structure and testing it against the simulated earthquake. Thus, it would appear that the added benefit of engaging in verbal interactions that are characteristic of the *Accept BDM* pathway is that the initial design ideas become more refined and consequently "better" which results in the building of more structures that are stable enough to withstand the simulated earthquake. Success on the earthquake task appears to be related to the articulation and modification of design ideas within the group. However, results suggest that it is not the number of proposed modifications that matters, as these were essentially the same for both high and low performing groups. Rather, it is likely that the quality of the proposed modifications is what makes a difference. Additional analyses will need to be done to explore this possibility.

If we look at the verbal interaction patterns of teams that were considered high performing based on their percentage of successful designs, a somewhat similar pattern emerges as is seen when teams are considered high performing based on the highest successful structure they built. While teams that had a high percentage of successful designs verbal interactions did not use the *Accept B* pathway more frequently than did the low performing teams, they did have marginally significantly more successful designs. This suggests that regardless of the definition of "high performing" that is used, verbal interactions that follow the *Accept B* pathway lead to success on the Earthquake task, and in turn, are characteristic of productive collaboration. This pathway suggests that the process of "rapid-prototyping," building and testing design ideas with minimal discussion of those ideas can be effective and lead to success in on this design task. Based on this finding, perhaps a helpful strategy would be to encourage teams employ this "rapid prototyping" process, to take advantage of the unlimited opportunities for testing. In addition, the *Accept BDM* verbal interaction pathway appears to be one that is characteristic of verbal interaction that leads to productive collaboration. Teams considered high performing due to their percentage of successful designs, had significantly more successful designs when this pathway was used, than did the low performing teams. Again, the number of proposed modifications made does not appear to make a difference to the effectiveness of this verbal interaction pathway to lead to productive collaboration.

Surprisingly, verbal interactions that are characteristic of the *Discussion* pathway do not appear to foster productive collaboration or success on the Earthquake task. The only significant difference between the *Discussion* pathway and the *Accept BDM* pathway, is that prior to building a structure the design ideas are discussed amongst the group members. Once the initial idea has been discussed and proposed modifications considered, building of the structure commences until it is ready to be tested against the simulated earthquake. Overall, teams rarely engaged in verbal interactions that followed the *Discussion* pathway. However, when teams' verbal interactions did follow the *Discussion* pathway, those that were considered high performing based on their percentage of successful designs, had marginally significantly more successful designs than did the low performing teams. This was not the case for teams that were considered high performing based on the highest successful structure built. Because all teams used the *Discussion* pathway minimally it is difficult to assess whether verbal interactions of this type do indeed lead to productive collaboration.

The results of this study lend support to the claims in the literature that successful verbal interaction while collaborating involves: (a) clear presentation of ideas, (b) elaboration on ideas by group members, and (c) reasoning and evaluation of ideas collaboratively (Hoek & Seegers, 2005; Mercer, 1996). Teams that were high performing engaged in verbal interactions that allowed them to engage with each other's ideas, refining and improving upon them, until an ideal solution was reached. In some regards, the Earthquake task was an ideal task for promoting productive collaboration, because the physical component of the task allowed individuals to make their ideas clearly visible to their team members simply by building their idea. Barron (2003) highlights the creation of a "joint problem-solving space as being particularly important for achieving productive collaboration. It can be argued that the Earthquake task in particular, and design-based learning tasks more generally, are particularly effective at facilitating this process for students because it requires the creation of an artifact that represents the thinking of the team. Teams that were highly successful on the Earthquake task were likely able to take advantage of being able to make their thinking "visible" through their design. This explanation may also help account for why teams did not make extensive use, and had less success with, verbal interactions that followed the *Discussion* pathway.

## Future Research

While it was already noted that the average number of proposed modifications did not differ between high and low performing teams for either the *Accept BDM* or *Discussion* pathways, it is possible that there was a qualitative difference in the modifications proposed that may have contributed to the overall success of the team. What types of modifications were proposed during the verbal interactions of the high performing teams and did the types of proposed modifications influence the outcome of the trials? Alternatively, it is possible that high performing teams began with "better" ideas, and thus discussion or modification of those ideas was not necessary. These questions will need to be answered to provide greater understanding of the verbal interactions that foster productive collaborations.

Finally, in an effort to further understand collaboration within teams, an additional future study could look more closely at specific team participants who may have dominated a team's overall performance by verbalizing more than others. Examining the group dynamics, and the influence it has on the teams' overall performance will help further elucidate the characteristics of teams and their verbal interactions that lead to productive and successful collaborations.

## References

Apedoe, X., & Schunn, C. D. (2009). *Understanding how students solve novel design challenges.* Poster presentation, National Association for Research in Science Teaching, Garden Grove, CA.

Azmitia, M., & Crowley, K. (2001). The rhythms of scientific thinking: A study of collaboration in an earthquake microworld. In K. Crowley, C.D. Schunn, & T. Okada. (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings.* (pp. 51-82). Mahwah, NJ: Erlbaum

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*, 307-359.

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., & Bransford, J. D. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *The Journal of the Learning Sciences, 7*, 271-311.

Cognition and Technology Group at Vanderbilt. (1994). *The Jasper Project: Lessons in curriculum, instruction, assessment, and professional development.* Mahwah, NJ: Lawrence Erlbaum.

Collins, A., Hawkins, J., & Carver, S. M. (1991). A cognitive apprenticeship for disadvantaged students. In B. Means, C.Chelemer, & M.S. Knapp (Eds.) *Teaching advanced skills to at-risk students* (pp. 216-243). San Francisco, CA: Jossey-Bass.

Cohen, E.G. (1994). Restructuring the classoom: Conditions for productive small groups. *Review of Educational Research, 64*, 1-35.

Cross, N. (1994). *Engineering design methods: Strategies for product design* (2nd ed.). New York, NY: John Wiley & Sons.

Fortus, D., Krajcik, J.S., Dershimer, R.C., Marx, R.W., & Mamlok-Naaman, R. (2005). Design-based science and real-world problem-solving. *International Journal of Science Education, 27*, 855-879.

Goel, V. & Pirolli, P. (1992). The structure of design problem space. *Cognitive Science, 16*, 395-429.

Hoek, D.J., & Seegers, G. (2005). Effects of instruction on verbal interactions during collaborative problem solving. *Learning Environments Research, 8*, 19-39.

Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The Elementary School Journal*, *94*, 483-497.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., & Soloway, E. (1997). Enacting project-based Science. *The Elementary School Journal*, *97*, 341-358.

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*, 359-377.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Reynolds, B., Mehalik, M. M., Lovell, M. R., & Schunn, C. D. (2009). Increasing student awareness of and interest in engineering as a career option through design-based learning. *International Journal of Engineering Education, 25*(1), 788-798.

Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understandings of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences*, *4*, 131-166.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.) *Handbook of Educational Psychology* (pp. 841-873). New York: Prentice Hall.

## Author Note

This material is based upon work supported by the National Science Foundation under EEC-0808675. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

# Implementing a Lesson Plan Vs. Attending to Student Inquiry: The Struggle of a Student-Teacher During Teaching Science

L. T. Louca, M. Santis & D. Tzialli, Department of Education Sciences, European University Cyprus,
6, Diogenous Str., Engomi, 1516 Lefkosia, Cyprus
Email: Louca.L@cytanet.com.cy, M.Santis@euc.ac.cy, D.Tzialli@euc.ac.cy

**Abstract**: Despite calls for student-centered, inquiry-based instruction in science, science teacher preparation remains mostly teacher-centered, with the underlying assumption that novice teachers need to form a teaching identity before attending to their students' inquiry. In this paper, we use the idea of framing to analyze a 42-minute science lesson of a senior kindergarten student-teacher. Findings suggest that the student-teacher struggled for balance between teaching science as implementing a lesson plan, and as attending to her students' inquiry. We use this evidence to suggest that novice teachers can attend to students' inquiry as early as in their student-teaching experience, which suggests additional pressure on the need for preparation in teaching science. Thus, the role of science methods courses should be to help students understand the different interpretations of teaching within the different frames and provide them with strategies for entering more productive frames during teaching.

## Introduction

Despite decades of calls for promoting students' inquiry in science in grades K-12 (NRC, 2000; 2007), this agenda has been slow to become established in instructional practice (Minstrell & van Zee, 2000; Osborne et al., 2004), possibly for a number of reasons. First, despite a wide consensus regarding the importance of inquiry in science learning, the education community has yet to agree on precisely *what* is important in scientific inquiry (Anderson et al, 2000). For many, inquiry is a method for learning science "content," while for others, inquiry is a part of science and an objective in itself. Secondly, there is no agreement regarding what "productive" inquiry should include, especially in the early grades. Answers have varied from general appeals for "messing about" to more specific targets for developing "concrete" abilities such as controlling variables (e.g., Metz, 1995).

In contrast to the tangible and more straightforward objectives of traditional content, when considering these ambiguities, the difficulty of sustaining instructional attention to student inquiry is understandable (Hammer, 1995). Regardless of the particular account of children's inquiry, there always exists the challenge of diagnosing student progress in any classroom situation (e.g., Goodwin, 1994). Developing such diagnostic abilities for identifying and responding to students' scientific inquiry depends largely on teachers' professional development (both pre- and in-service) in teaching science. While a number of studies have looked into how teachers in primary and secondary education teach science (e.g., Kuiper, 1995), little is known about how kindergarten teachers implement the science aspects of the curriculum (e.g., Kallery & Psillos, 2002). To make progress in promoting student inquiry especially in early grades, science education needs to develop a better understanding of how teachers perceive and respond to student inquiry in classroom settings and what struggles they encounter to implement inquiry-based teaching.

The purpose of this case study is to contribute towards this understanding and development, by proposing and applying a specific methodological approach which helps to describe how a senior student-teacher conceived her role in the class and how her behavior reflected a struggle to balance between teaching science as implementing a lesson plan she had developed and received approval for by her tutor, in contrast to attending to her students' scientific inquiry. We use the idea of framing (Goffman, 1974; Tannen, 1993; Schank, 1990) to analyze a 42-minute science lesson of this student-teacher. In doing so, we seek to show how using the idea of "framing" in the context of analyzing student-teaching may prove useful in understanding the process of novice teachers' professional development and in identifying areas in which teacher education may support this. We also argue that it is not unrealistic to expect novice teachers to be able to attend to and respond to students' inquiry during real-time teaching.

## Theoretical framework

### Emphasis of science education in inquiry-based, student-centered approach

Following recent changes and reforms in curricula, science has become established as part of the early primary curriculum (The Curriculum Guidance for the Foundation Stage QCA/DfEE, 2000). The National Research Council (2007) strongly recommended the development of an approach for teaching science in kindergarten through eighth grade, in which science is viewed as providing opportunities for both learning and development, or to lay down the foundations in preparation for future learning in science (Kamii & DeVrie, 1993). In this context, considerable attention is given to children's own explorations and inquiry (QCA/DfEE 2000), with a number of researchers emphasizing the importance of engaging young children in hands-on science experiences which support the development of an early interest and knowledge base in meaningful scientific themes, and which provide children with an introduction to and support for developing science inquiry skills (e.g., Chen & McNamee, 2007; Kamii & DeVries, 1993).

Current emphasis in science education highlights among other things the support of student inquiry (NRC, 2000; 2007), promoting it as the central strategy for teaching science. The National Science Education Standards (NRC, 1996) has provided a definition modeled after the work of scientists where scientific inquiry refers to the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work. Inquiry also refers to the activities of students in which they develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world. Numerous other definitions can be found in the literature, which vary in their emphasis (e.g., Barman, 2002; Flick, 2002; Crawford, 2006). Taking all these into account, for the purpose of this paper, we take student inquiry to mean the pursuit of causal, coherent explanations of natural phenomena (Hammer, 2004) which may take many forms, both experimental and theoretical. Regardless of the form, the instructional agenda is to help students learn to engage in that pursuit for themselves. In this view, science inquiry supports the development of problem solving, communication and thinking abilities as students pose questions about the natural world and seek the evidence to answer them (NRC, 2000).

## Preparing teachers for inquiry-based teaching in science: approaches and challenges

Despite NRC's (2000; 2007) current emphasis on "student-centered" instruction in science education, only a few programs have been designed to support pre-service teachers in developing their knowledge about scientific content and inquiry (e.g., McDevitt et al., 1995). On the whole, teacher preparation for teaching science remains largely teacher-centered. Science methods courses focus on the things teachers (should) do, on instructional methods (Tobin & Fraser, 1990), management strategies (Feiman-Nemser & Parker 1992), questioning skills (Fleer & Hardy, 2001), evaluation procedures (Jarvis et al., 2001), and planning processes (Lenton & Turner, 1999), putting much of the emphasis of teacher preparation programs on helping teachers form a teaching identity (e.g., Freese, 2006).

The teacher-centered approach of preparing pre-service teachers to teach science has been strongly influenced by stage-based accounts of teacher development (Kagan, 1992). Teacher developmental stages refer to the stages through which teachers progressively gain professional knowledge, abilities and beliefs (Nimmo, 1994). This approach proposes that during initial stages of their development, novice teachers need to clarify and construct their self-image as a teacher by developing routines related to classroom management and instruction. Beginning teachers' concerns tend to focus primarily on self and self-image as a teacher, rather than on students' learning. Consequently, the level of effort and skill required by beginning teachers to teach science using inquiry-based approaches may not be achievable for many. Therefore the tendency in these instances is to revert to authoritarian, teacher-directed approaches (Harlen, 1996). In this context, activities or units that work are seen to be effective partly because they engage students, in the sense that they keep students interested. Although this approach results in teachers focusing on their own behavior before they can attend to student learning, it is seen as a crucial step early in their professional development (Kagan, 1992). Once they have "resolved" an image of themselves, novice teachers can then shift their attention to attending to and responding to their students' inquiry. Thus, the challenge to integrate aspects of teaching science as inquiry, into their planning and instruction maybe frustrating for novice teachers.

Conversely, Crawford (1999) suggests that it is not only possible, but it is also realistic to expect that at least some novice teachers design and carry out inquiry-based instruction. Novice teachers can be capable of articulating an emerging knowledge of teaching science as inquiry, and they may espouse philosophies aligned with this kind of pedagogy (Crawford & Lunetta, 2002; Windschitl, 2003), but they may be incapable of, or unwilling, for various reasons, to enact teaching science as inquiry in their classroom (McGinnis, Parker, & Graeber, 2004; Newman et al., 2004).

Levin, Hammer and Coffey (2009) provide an extensive review of the challenges to stage-based accounts of teacher development, indicating that this view may be misleading and steer teacher education in science in unproductive directions. They suggest that in a number of studies, pre-service teachers were able to reflect on several issues related to the content of teaching (Grossman, 1992), and to attend to the substance of their students' thinking (Davis, 2006), although Davis questions pre-service teachers' abilities to reflect-in-action. Despite stage-based accounts implying that after having developed classroom routines teachers focus their attention to student reasoning, Zeichner and Gore (1990) showed that novice teachers rapidly shift from progressive, student-centered attitudes formed during pre-service to traditional, teacher-centered approaches when confronted with the realities of the workplace. Furthermore, studies showed that teachers often become satisfied with their teacher-centered approach to teaching and are less likely to question their chosen routines (Grossman, 1992). Finally, Olsen (2007) indicates that such stage-based models focus on implementation and not necessarily on decision-making processes that take place during teaching and learning, and that such work has been primarily conducted with in-service teachers.

We agree with these challenges and we highlight that the ability to attend to students' inquiry is one of the critical aspects of the pedagogical content knowledge (PCK) that novice teachers need to develop as part of their pre-service preparation (e.g., Davis & Smithey, 2008). Even though helping pre-service teachers develop rich pedagogical content knowledge prior to substantial teaching experience is not an easy task (van Driel, De Jong, & Verloop, 2002). We agree with Davis & Smithey's suggestion that it is possible to provide pre-service teachers with enough experiences to prepare them to develop "PCK readiness" – that is the requirements for developing abilities to attend to their students' inquiry. Davis & Smithey use the following analogy to illustrate the idea of "PCK readiness": Kindergarten teachers spend most of the school year helping students develop "reading readiness" i.e., the skills students need to develop before they can read. None of these skills (how to hold a book, which direction to turn the pages, the letters and their sounds) is technically "reading," but every reader needs these skills. Similarly, they suggest that pre-service teachers can learn about the content, how to represent this to learners, and common ideas learners bring to the science class. Even if this initial knowledge is in pieces, this forms the building blocks for well-developed and usable PCK.

### The idea of framing and how we transfer it to student-teaching

Though substantial gains have been made in understanding teachers' professional knowledge development (e.g., Munby et al., 2001; Calderhead, 1996; Borko & Putnam, 1996), little is known about the difficulties novice teachers encountered during teaching science as inquiry (Anderson, 2002), raising questions about how to support learning and to enact teaching science as inquiry (Newman et al., 2004; Windschitl, 2003).

To contribute towards this end, we have adopted the term "framing" based on previous work in sociology (Goffman, 1974), sociolinguistics (Tannen, 1993), and cognitive science (Schank, 1990). Framing is used to describe the idea that people use knowledge from past experiences to make sense of what is going on in situations they perceive to be similar. In this sense, individuals are accomplished at attending to what is happening around them, searching for signals that indicate the type of activity and making alterations to their behavior when it appears appropriate (MacLachlan & Reid 1994). For instance, when entering an unfamiliar restaurant people immediately interpret what kind of restaurant it is based on activities around them: Are customers waiting for a waiter to take their order or are they standing at a counter ordering food? In this manner framing influences one's expectations about what is going on as well as their interpretations of what they see or expect to happen in particular situations, so influencing the way they act.

We can parallel the manner people frame daily activities with the way pre-service teachers frame their role in science classrooms. There are many aspects to framing, including how to manage teaching time, maintain student control and introduce new activities. Additionally, the consideration of whether or how teachers respond to their students' inquiry may also be an aspect of framing. Their understanding of successful teaching should include the issue of responding to students' inquiry or not, and in which ways this is done. This perspective may help to better understand student-teachers' enactment of teaching during the early phases of their careers.

Our purpose in this paper was to describe how the a senior student-teacher conceived her role in the class and how her behavior reflected a struggle to balance between teaching science as implementing a prepared lesson plan, and as attending to her students' emergent inquiry, seeking to identify different ways of teacher acting during teaching. While it is inconceivable to focus only on implementing a lesson plan with total disregard of student inquiry in a student-centered setting, it is also "unacceptable" for a student-teacher to ignore an informed lesson plan. This dilemma has implications for what is seen as the role of science teaching preparation and how educators support their student teachers' struggle to balance both criteria. In this paper, we seek to show that the student-teacher was able to attend to her students' inquiry during her teaching early in her career, despite stage-based accounts of teacher development.

## Methodology

This is an interpretive case study focusing on a single science lesson of a senior kindergarten student-teacher. We recorded and transcribed this lesson in March 2009 as part of the student-teachers' requirements to have five of their science lessons videotaped, which they subsequently reviewed and discussed with the faculty teaching science methods, their teaching supervisors and their peers.

For her last semester of student-teaching (Spring 2009) Mary (the study's participant) was assigned to an urban public kindergarten school and carried out her nine-week student-teaching practicum in a class of 17 students (age ranging: 4.5 - 5.5 years old). The lesson we analyzed for this paper was her first science lesson during this student-teaching phase and concerned the solubility of various substances in water. We chose to analyze this specific lesson because it was easy to identify instances in which Mary followed her lesson plan or not, as she had an elaborated version of her lesson plan and when following it she would literally read from it.

For this paper, we used the lesson video and its transcript as our primary data source. Using Videograph software, we coded all the episodes during which Mary talked during whole class discussions. The coding was based on whether she was following her lesson plan, or making "teaching moves" for which she had not planned. We used four codes for this analysis where Mary: (1) read directly from her lesson plan, (2) did not read from her lesson plan, (3) followed an open-ended discussion activity that she had planned for (that is she did not read from her lesson plan, but she was following a pre-planned activity) and (4) students talked. The coding was carried out by the second and third authors (Cohen's Kappa=0,945) and the resulted differences were resolved through discussion.

When agreement was reached, in order to have participant check of our interpretations of the findings, the first author carried out an open-ended interview with Mary, during which the video of the lesson was reviewed, pausing it at various points that had been selected based on the findings of the analysis. At those points, he asked Mary about her thinking and purpose for specific "teaching moves". In some cases, she requested to watch the episode under discussion once more. Eventually, she was able to identify the elements of the students' thinking to which she was responding. In this way, we verified or made alterations in our interpretations of the findings of the code-based analysis.

## Findings

For about 14 minutes of the lesson (out of the total 42 of the entire lesson – 33.3%) Mary read directly from her lesson plan, obviously following an elaborately prepared plan. Out of the total 483 utterances coded, this time corresponded to 86 utterances (17.8%). We refer to these instances as working within an *implementing-a-lesson-plan frame* (Frame 1). Conversely, for about nine minutes, (21.42% of the total lesson time) Mary spoke spontaneously without searching to read what to say. This time corresponded to 125 utterances (25.87%). Within these nine minutes of spontaneous talk, we identified two distinct sub-frames where Mary either a) responded directly to her students' thinking or conversational contribution (*attending-to student-inquiry-frame* – Frame 2) or b) she was following an open-ended discussion with her students within her lesson plan's general structure (*lesson-plan free-talk frame* – Frame 3). The second frame consisted of 49 utterances (10.14%) and lasted for about 4 minutes (9.5% of the total time), and the third frame consisted of 76 utterances (15.73%) and went on for about 5 minutes (12.5% of the total time). The remaining 217 utterances corresponded to student talk in the conversations, while the time unaccounted for was taken up by the children's experimentation (about 9 minutes). Figure 1 represents a timeline graph of the lesson's discourse. From this it is clear that, while there are instances where Mary worked primarily within one of the three frames identified (e.g., utterances 0-13; 81-145; 416-449), there are also instances when she shifted to and from between the frames over a short period of time (e.g., 14-80; 140-231; 350-455). The analysis of the transcript and the subsequent interview with Mary, allowed us to define the characteristics of the frames identified. Below we describe in detail those characteristics for each of the frames we identified.

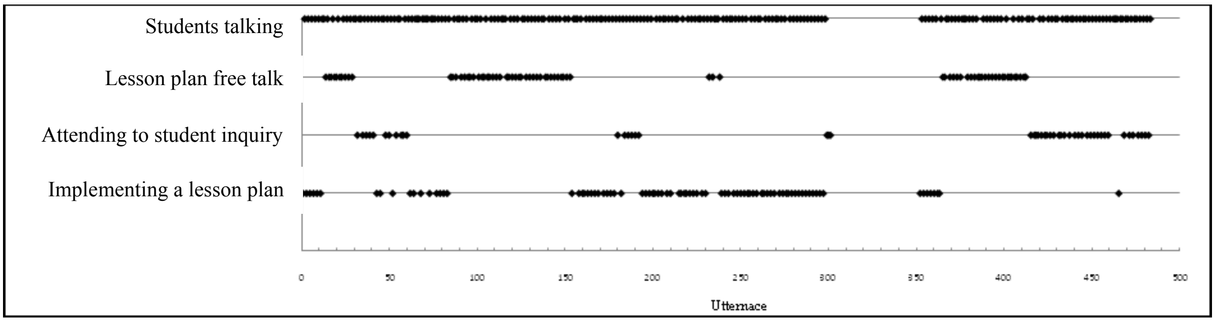



Figure 1. Timeline graph of the student-teacher's discourse

### Working within the implementing-a-lesson-plan frame

For most of her teaching time, Mary worked within the implementing-a-lesson-plan frame. She began her lesson within this frame, and in almost all cases returned to this frame after having shifted to the other frames. From the interview, it was clear that Mary was very conscious of the fact that the lesson plan she had prepared included prescribed activities from the teacher's official guide for teaching science in kindergarten, and for which her tutor had given approval thus, she felt that "these are the steps we have been taught to follow". Her sense of urgency to fulfill each outlined step was also apparent as she proceeded with the sequence of activities, paying little attention to her students' responses and understanding. Her questioning, as she turned from one student to the next, was characterized by the search for the correct, predefined answer. The conversational exchanges that took place in this frame lasted longer than the other two frames, with each of her utterances lasting for an average of 10s – (the duration in frame 2 was about 5 seconds and about 4 seconds in frame 3.) Consequently, her questioning sessions during this frame labored for as long as it took to find the answer required. One of the most salient aspects of this frame was the jarring changes from one activity to the next when Mary felt she 'had' to move on.

In the example below we see Mary cutting short a discussion in order to move on to the next activity. This discussion, in which students were articulating their hypotheses relating to what had happened to the sugar in the water, was carried out in attending-to-student-inquiry frame. In this excerpt, Mary reverted back to the lesson plan's next activity (frame 1) without linking it to her students' on-going discussion, giving the impression that it was the lesson plan, not Mary 'running' the lesson.

*Mary:* Harry says the sugar fell into the water, so what might have happened to the sugar? [as it cannot be seen]
*Student 1:* It turned into water
Mary: What do you mean it turned into water?
*Student 2*: It melted
*Mary*: It may have melted…
*Student 1*: It became sugar
*Mary*: OK, now I am going to give you each something, OK? You can touch it, smell it, but you mustn't taste it, [hands out substances with which the children will experiment]

## Working within the attending-to-student-inquiry frame

Our analysis showed that Mary shifted from following her lesson plan to directly responding to her students' thinking, and in this attending-to-student-inquiry frame, she did not read from her lesson plan. Rather, she attended and responded to her students' thinking, even if this led her outside of what she had planned. In these instances, she spontaneously provided her students with some feedback, comment or conversational direction that was not part of her lesson script. This frame usually lasted for short bursts of dialogue when Mary responded directly to her students' inquiry and then she quickly reverted to frame 1. For instance, after the children's various predictions of what would happen to the substances in water, during which Mary followed her lesson plan, the different children's predictions caused Mary to abandon her lesson script.

*Mary*: How can we be sure what is right? I see we have some disagreements… how will we know for sure if our materials will dissolve or not? What should we 'scientists' do?
*Student 3*: We can sit at our tables and see…
*Mary*: See what? Do what at our tables?
*Student 3*: Scientific things!
*Mary*: Scientific things! Well done, and what do we call these scientific things?
*Student 4*: Experiments!
*Mary*: Well done Nancy! We're going to do an experiment!

After this short discussion (frame 2) , Mary shifted back to following her prescribed lesson plan (frame 1) and brought out a chart for noting down the predictions prior the experiments. We suggest that the relatively brief duration of these exchanges in frame 2 (an average of 5 seconds per teacher utterance) was due to two factors. First, the fact that the dialogue during this frame was not part of the "official script" seemed to cause Mary anxiety and a desire to return back to the security of her lesson plan. Second, it became apparent during the interview that Mary somehow underestimated the real value of these activities regarding them as only perfunctory steps in the process towards understanding the concept of dissolving. Nevertheless, during this frame there was cohesion not only within the frame as Mary's questioning directly addressed the children's conversational contributions, but also as she moved from this frame to a different one. These moves were characterized by a natural flow which did not disrupt the rhythm of the lesson nor disorientate the children.

## Working within the lesson-plan free-talk frame

During analysis we identified a third, somewhat intermediate, frame to the previous two. This frame was observed (1) during the discussion of the problem presented (where was the sugar which had fallen into the water?), (2) while the children identified the materials with which they would experiment, and (3) immediately following their experiments, when they announced their results to the rest of the class. In this lesson-plan free-talk frame, Mary also responded to the children without reference to her lesson plan, but on these occasions she was clearly working within a particular activity she had noted previously as a sub-heading but had not scripted for in the lesson plan. She responded less specifically to the children's conversational contributions compared to the second frame, often reflecting their answers back at them when these were wrong (e.g., "is it sugar?") or repeating their correct answers, meanwhile using more closed questions than when working within the attending-to-student-inquiry frame. This frame always followed on from the implementing-a-lesson-plan frame and it was apparent that Mary was aware of the pedagogic objective for these activities which, to some extent restrained her responses to students' inquiry in her attempt to fulfill the objective. On two occasions, when she moved from this third frame to the attending-to-student-inquiry frame, this occurred when the children encountered difficulties in responding appropriately to Mary's more restricted questioning and Mary seemed to instinctively and effortlessly change her framing to accommodate for this factor and respond directly to the children's comments, building on these comments to guide the children in their discovery.

The following is a short excerpt in which the children announced their conflicting results for whether flour dissolved in water or not. In this excerpt, Mary shifted from working in the lesson-plan free-talk frame (during which she guided students to present their results) to attending-to-student-inquiry frame.

*Mary*: Let's look at your results for flour. What happened to the flour? Did it dissolve or not dissolve?
*Student 5*: It dissolved!
*Student 2*: No, it didn't dissolve!
*Mary*: What did you discover? Let me come and see [what they had noted on their charts]
*Student 1*: We discovered that sugar….
*Mary*: But what happened to the flour now. […] Did it [the flour] dissolve or not?
*Student 1*: It dissolved.
*Mary*: Has another team noted something different? Do you agree?
*Student 4*: Andy's team discovered that it didn't dissolve.

While working within a lesson-plan free-talk frame for some time, Mary "saw" that at least two groups of students had different results for whether the flour dissolved in the water or not. A possibility is that this disagreement is related to how one defines solubility (e.g., something is considered dissolved if it is not visible any more, or when one cannot identify two separate substances). While Mary did not clarify this definition, she addressed the disagreement between the groups, because, as she indicated in the interview, students acting as scientists need to reach consensus about an experimental result. To do that, she shifted to the attending-to-student-inquiry frame. Mary also indicted that when she reverted back to following-the-lesson-plan frame, it was specifically to move to the next activity in the lesson either guided by time restraints or believing that she had fully addressed the children's questions. These moves back to the following-the-lesson-plan frame were marked by abrupt changes in the discussion's direction and lesson activity.

### Themes emerged across the three frames

Despite the differences between the three frames that we described above, we have identified at least two emerging themes relevant to all the frames. These include Mary's need for "teaching security" and her role in the classroom, which we describe below.

Mary explained that the lesson plan offered her security by knowing "where [she was] in the lesson and how to proceed." She indicated that "I feel secure that I know now I do this and then I do that." This had an impact on the duration of the various activities: During the lesson implementation, Mary felt justified to allow e.g., the questioning to last for a long time, until the children gave the 'correct' answer, resulting in her paying little attention to alternative answers offered. Conversely, she was aware of the time pressure during the frames outside of the lesson plan stating emphatically, "the discussion flowed and the lesson progressed but it took too long." Her view that these activities did not contribute substantially to the learning added to her need to return to the safety and validity of the lesson plan, and was evident in her abrupt jumps from the lesson-plan free-talk back to the implementing-the-lesson-plan frame.

Mary also indicated that while working within the implementing-a-lesson-plan frame she "just let the lesson plan roll and didn't respond to them [her students]", supporting our own findings that in this frame Mary "automatically" moved from one activity to the next. She conceived her role as following an approved plan, and thus her main concerns were to finish the activities culminating in the correct answer, though she described her role as that of a guide, providing the stimuli with which to develop the children's knowledge and discovery in the direction she saw as relevant. In contrast, while working within the attending-to student-inquiry-frame or the lesson-plan free-talk frame, she indicated that the children had made comments she had not predicted and thus responded intuitively addressing the issues raised. Likewise, when asked why she had not scripted for all the activities in the lesson plan, she said that during these activities "I knew [what needed to be done]".

## Discussion

Our purpose in this study was to describe in detail a science lesson of a senior kindergarten student-teacher in order to gain better understanding of how she conceived her role in the class and how her behavior reflected a struggle to balance between teaching science as implementing a lesson plan she had developed and received approval for by her tutor, in contrast to attending to her students' scientific inquiry.

Our first contention is that our findings show that the idea of "framing" (Goffman, 1974; Tannen, 1993; Schank, 1990) as a methodological approach used for analyzing teaching may prove useful in understanding the processes that novice teachers go through during their first attempts to teach. Framing is used to describe the idea that people are accomplished at attending to what is happening around them by searching for signals that indicate the type of activity and making alterations to their behavior when it appears appropriate (MacLachlan & Reid 1994), and by using past experiences to make sense of what is going on in situations they perceive to be similar. This helped us define different ways of teacher acting during teaching and their characteristics and so identify several issues that caused Mary's struggles during student-teaching in science.

The need to follow an approved lesson plan, to feel confidence in what she was doing, the restriction of the time, and her understanding of her role as a teacher in the classroom were some of the factors that guided Mary's behavior during the lesson we analyzed.

Second, we argue that it is not unrealistic to expect novice teachers to be able to attend to their students' inquiry during actual teaching. Despite views suggesting that novice teaches may be incapable of enacting teaching science as inquiry (e.g., McGinnis, Parker, & Graeber, 2004), our findings suggest that the student-teacher in this study was able to attend to her students' inquiry at times and successfully integrate aspects of teaching science as inquiry into her teaching. Of course, we do not suggest that she was expert in teaching science as inquiry, in fact, there are several respects in which it is evident she was not. Our conclusion, then, is that novice teachers come to their student-teaching with some possibly nascent abilities for inquiry teaching, that they may invoke these spontaneously, directly in line with the abilities that teacher educators have described as important to impart. That Mary actually shifted frames, suggests having the ability to make these transitions. From the interview it was clear that these shifts were not always done consciously, but intuitively.

Agreeing with challenges to stage-based accounts of teacher development (e.g., Levin, Hammer & Coffey, 2009; Grossman, 1992; Davis, 2006; Sherin, 2004), we suggest that an important role of undergraduate methods courses is to help students make clear distinctions between alternative teaching frames, understand the different interpretations of teaching within each frame and the characteristics of the teaching resulting from each. At the same time it is important to help pre-service teachers develop strategies both for (a) entering the attending-to-student-inquiry frames during early stages of their careers, and (b) working within that frame more reliably. This will help pre-service teachers to develop "PCK readiness" (Davis & Smithey, 2008) which they can then use for developing abilities to attend to their students' inquiry.

## References


Anderson, R. D., & Mitchener, C. P. (1994). Research on science teacher education. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 3–44). New York: Macmillan.

Anderson, L. M., Smith, D. C., & Peasley, K. (2000). Integrating learner and learning concerns:Prospective elementary science teachers' paths and progress. *Teaching and Teacher Education*,16(5–6), 547–574.

Anderson, R.D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, 13, 1-12.

Barman, C.R. (2002). How do you define inquiry? *Science and Children*, 26, 8–9.

Borko, H. & Putnam, R.T. (1996). Learning to teach. In Berliner, D.C. & Calfee, R.C. (Eds.), *Handbook of educational psychology* (pp. 709-725). New York: Macmillan.

Calderhead, J. (1996). Teachers: Beliefs and knowledge. In Berliner, D.C. & Calfee, R.C. (Eds.), *Handbook of educational psychology* (pp. 709-725). New York: Macmillan.

Chen, J., & McNamee, G. D. (2007). *Bridging: Assessment for teaching and learning in early childhood classrooms, preK–3.* Thousand Oaks, CA: Corwin Press

Crawford, B.A. (1999). Is it realistic to expect a preservice teacher to create an inquiry-based classroom? *Journal of Science Teacher Education*, 10, 175–194.

Crawford, B. A. (2006). Learning to Teach Science as Inquiry in the Rough and Tumble of Practice. Journal of Research in Science Teaching, 44, 613–642.

Crawford, B.A., & Lunetta, V. (2002). Promoting the development of a personal philosophy of teaching prospective secondary science teachers. *Pennsylvania Teacher Educator,* 1, 68–74.

Davis, E. A. (2006). Characterizing productive reflection among preservice elementary teachers: Seeing what matters. *Teaching and Teacher Education*, 22, 281-301.

Davis, E. A., & Smithey, J. (2008). Beginning Teachers Moving Toward Effective Elementary Science Teaching. Science Education, 93, 45-770.

Flick, L. (2002). *Inquiry as cognitive process.* Paper presented at the annual meeting of the NRST, Philadelphia.

Feiman-Nemser, S., & Parker, M. (1992). *Mentoring in context*: *A comparison of two U.S. programs for beginning teachers*. NCRTL Special Report. East Lansing, MI: NCRTL

Fleer, M., & Hardy, T. (2001). *Science for children*. Sydney, Australia: Prentice Hall.

Freese, A. R. (2006). Reframing one's teaching: Discovering our teacher selves through reflection and inquiry. *Teaching and Teacher Education*, 22, 100-119.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. New York: Harper & Row.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96, 606–633.

Grossman, P. L. (1992). Why models matter: An alternate view on professional growth in teaching. *Review of Educational Research*, 62, 171-179.

Hammer, D. (1995). Student Inquiry in a Physics Class Discussion. Cognition and Instruction, 13(3): 401-430.

Hammer, D. (2004). The variability of student reasoning, lectures 1-3. In E. Redish and M. Vicentini (Eds.), *Proceedings of the Enrico Fermi Summer School, Course CLVI* (pp. 279-340): Italian Physical Society.

Harlen, W. (1996) *The Teaching of Science* (2nd ed.) London: David Fulton.

Jarvis, T., McKeon, F., Coates, D., & Vause, J. (2001). Beyond generic mentoring: Helping trainee teachers to teach primary science. *Research in Science and Technological Education*, 19, 5–23.
Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62, 129-169.
Kallery, M. & Psillos D, (2002). What Happens in the Early Years Science Classroom? The Reality of Teachers' Curriculum Implementation Activities. *European Early Childhood Education Research Journal 10(2).*
Kamii, C. & DeVries, R. (1993). *Physical Knowledge in Preschool Education: implications of Piaget's theory.* New York, NY: Teachers College Press.
Keys, C. W., & Bryan, L. A (2001). Co-Constructing Inquiry-Based Science with Teachers: Essential Research for Lasting Reform. *Journal of Research in Science Teaching*, 38, 631-645.
Kuiper, W. (1995) *The Implementation of Context - and Activity Based- Science Education:Intention and Reality.* Paper presented at the annual meeting of NARST, San Francisco, CA
Lenton, G., & Turner, G. (1999). Student-teachers' grasp of science concepts. *The Journal for Science Education*, 81(295), 67–72.
Levin, D. M., Hammer, D. & Coffey, J. E. (2009). Novice Teachers' Attention to Student Thinking. *Journal of Teacher Education,* 60, 142-154.
MacLachlan, G., & Reid, I. (1994). *Framing and interpretation.* Carlton, Canada: Melbourne University Press.
McDevitt, T. M., Troyer, R., Ambrosio, A. L., Heikinen, H. W.& Warren, E. (1995). Evaluating prospective elementary teachers' understanding of science and mathematics in a model preservice program. *Journal of Research in Science Teaching*, *32*(7), 749-775.
McGinnis, R. Parker, P., & Graeber, A. (2004). A cultural perspective of the induction of five reform-minded beginning mathematics and science teachers. Journal of Research in Science Teaching, 41, 720–747.
Metz, K. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research*, 65, 93 – 127.
Minstrell, J. & van Zee, E. H. (Eds.). (2000). *Inquiring into inquiry learning and teaching in science.* Washington, DC: American Association for the Advancement of Science.
Munby, H., Russell, T. & Martin, A.K. (2001). Teachers_ knowledge and how it develops. In: Richardson, V. (Ed.), Handbook of research on teaching, AERA (pp. 877-904). Washington, DC.
National Research Council. (1996). National Science Education Standards. Washington DC: National Academy Press.
National Research Council. (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning*. Washington, DC: Com. on the Development of an Addendum to the National Science.
National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K*-8. Washington, DC: Committee on Science Learning, Kindergarten Through Eighth Grade.
National Science Education Standards (1996). *National Committee on Science Education Standards and Assessment*; National Research Council. Washington DC: National Academy Press.
Newman, W., Abell, S., Hubbard, P., McDonald, J., Otaala, J., & Martini, M. (2004). Dilemnas of teaching inquiry in elementary methods. *Journal of Science Teacher Education*, 15, 257–279.
Nimmo, G. (1994). The Idiosyncratic Nature of Beginning Teaching: Reaching Clearings by Different Paths. Paper presented at the Annual Meeting of the ATEA, Brisbane, Queensland, Australia.
Olsen, J. K. (2007). Preservice Teachers' Thinking Within a Research Based Framework: What Informs Decisions? *International Journal of Science and Mathematics Education.* 5, 49-83.
Osborne, J., Erduran, S. and Simon, S. (2004). Enhancing the quality of argumentation in school science. J. of Research in Science Teaching, 41, 994-1020.
QCA/DfEE (2000). *Curriculum guidance for the foundation stage.* London: Qualifications & Curriculum Authority.
Schank, R. (1990). *Tell me a story: A new look at real and artificial memory.* New York: Scribner
Sherin, M.G. & Han, S.Y. (2004).Teacher learning in the context of a video club. *Teaching & Teacher Education*, 20,163–183.
Tannen, D. (1993). *Framing in discourse*. New York: Oxford University Press.
Tobin, K., Kahle, J., & Fraser, B. (1990). Windows into science classrooms: Problems associated with higher-level cognitive learning. London: Falmer Press.
Tu, T. (2006). Preschool science environment: What is available in a preschool classroom? *Early Childhood Education Journal,* 33*,* 245–251.
van Driel, J., De Jong, O., & Verloop, N. (2002). The development of preservice chemistry teachers' pedagogical content knowledge. Science Education, 86, 572 – 590.
Windschitl, M. (2003). Inquiry projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Education*, 87, 112–143.
Zeichner, K., & Gore, J. (1990). Teacher socialization. In W. R. Houston, *Handbook of research on teacher education* (pp.329-348). New York: Macmillan.

# "I study features; believe me, I should know!" The mediational role of distributed expertise in the development of student authority

Jennifer M. Langer-Osuna, University of Miami
222 Merrick Building, Coral Gables, FL, 33124 USA, jlangerosuna@miami.edu
Randi A. Engle, University of California, Berkeley
4641 Tolman Hall, Berkeley, CA, 94720 USA, raengle@berkeley.edu

**Abstract:** This paper examines the development of student authority in a case of one student assigned the role of topic expert in a classroom that utilized distributed expertise as a participation structure during collaborative projects. Using video data of a heated student-led debate, we show how this student successfully positioned himself with greater authority than recognized adult experts, despite the fact that his evidence was often weak given classroom norms. He was able to do so for two reasons. First, he utilized his role as topic expert to position himself with authority and discredit others, including the recognized adult experts. Second, he drew on student allies that supported his position. We conclude with implications of this paper for how participation structures may mediate the development of powerful student identities, and function in concert with other interactional factors in the classroom.

## Introduction

Many math and science classrooms have shifted toward a sociocultural view of education, in which learning occurs through interactions among students, and between students and the teacher. A central feature of such classrooms is to have students work collaboratively in structured ways designed to support their engagement in disciplinary ideas and practices. Participation structures, such as using a "round robin" to share ideas (Langer-Osuna, 2007; Saxe et al., 2005), the jigsaw method for sharing expertise (Brown & Campione, 1994), or the use of student group roles (Cohen & Lotan, 1995; Herrenkohl & Guerra, 1998) afford students opportunities to think through and debate ideas, to judge ideas as reasonable, and to gain a sense of ownership over their learning. Participation structures are also designed to support more equitable engagement in collaborative tasks (Cohen & Lotan, 1997). If each student is expected to contribute ideas in a structured fashion, then it is less likely students will become privileged in or marginalized from group discussions.

Recent studies, however, have shown that the goals of equitable, collaborative learning are difficult to accomplish even with such participation structures because students tend to construct inequitable relations of authority among themselves (Bianchini, 1999; Kurth, Anderson, & Palinscar, 2002). For instance, Esmonde and Langer-Osuna (2009) found that relational power among students—including social authority such as friendship and popularity, and academic authority such as being considered smart—still played a large role in how students negotiated group discussions despite classroom norms valuing multiple competencies. Engle, McKinney de Royston, Langer-Osuna, Bergan, and Mazzei (2007) found that some students illegitimately gained more influence and authority over others in a discussion in part by differentially making use of tactics to gain greater access to the discussion despite classroom norms designed to promote equitable turn-taking.

Understanding how students become positioned with authority is important to the goal of supporting productive student-led discussions (Engle & Conant, 2002; Gresalfi, Martin, Hand, & Greeno, 2009). Students positioned with authority participate more frequently in small groups, are more able to gain access to and hold the conversational floor, decide what is correct, be seen as contributing more meritorious ideas, and become more influential than students perceived as having less social or academic authority (Cohen, 1997; Inglis & Mejia-Ramos, 2009; Langer-Osuna, 2009). Being positioned with authority supports engagement because the student is in a position to judge what is correct or incorrect, and to initiate or close down group discussion (Esmonde & Langer-Osuna, 2009). Such privileged engagement and influence can occur not only in classroom using participation structures, but even when the students positioned with authority make claims that are not as objectively meritorious as those made by others (Engle, et al., 2007; Kurth, Anderson, & Palinscar, 2002).

This paper focuses on how a particular participation structure, the role of being a local expert mediated the assigned student's position of authority in coordination with his and other students' actions. The classroom used a participation structure called *distributed expertise,* in which students were assigned particular topics to research and become experts in, and then teach what they have learned to other students (Brown et al., 1993). In particular, we explain how this student, Brian, ultimately became positioned with authority because: (a) he utilized his role as features expert in order to position himself with greater authority than other students and even adult professionals; and (b) he made use of potential allies who took up his position of authority.

## Data Collection and Analytic Strategies

In classrooms that utilize distributed expertise, students are treated like researchers who gather, analyze, and share information about a particular topic, gradually developing expertise in that specialty area (Brown et al., 1993). Our paper is situated in one such classroom, in which students were researching different endangered species. Students in this classroom established their expertise when they showed that they could provide a legitimate answer to any reasonable question about their topics (Engle & Conant, 2002).

We analyze a case of nine fifth graders who engaged in a student-led debate (Engle & Conant, 2002; Engle, et al., 2007) that arose unexpectedly the day after a field trip to Marine World, then a science learning center. During the orca (i.e., killer whale) show one of the trainers stated that orcas were not whales, but rather the largest member of the dolphin family. Soon before going to Marine World and after a long struggle, this particular group of fifth graders had finally proven themselves to have sufficient expertise about whales to be allowed to create a bulletin board about their research for the rest of the school. As part of that, Brian had been established as the local expert on the anatomical features of whales. Thus, when the Marine World trainers mentioned the orca's large dorsal fin as being relevant to their assertion that orcas were not whales, this was taken up as a direct challenge to the group's, and especially Brian's, hard-fought authority. In the face of what is typically a greater source of authority – adult professionals – Brian stood firm on his position as an "expert" on whale features. Thus he sought to position himself as *more credible* than the adult professionals.

We view the case of Brian as a revelatory case (Yin, 2001) as it is more dramatic than typical student-led discussions, thereby highlighting how issues of authority are negotiated vis-à-vis student roles in interaction. Indeed, the student-led debate was loud and intense, and Brian successfully utilized his role as features expert to convince all but one student that he was a more credible source of information than the Marine World trainers and, later, the author of a science book about whales. The dramatic nature of this debate illuminates the ways in which such participation structures as distributed expertise can mediate not only the participation of a particular student in a given role, but also the interactions among students, and how students get positioned as a result. The processes of mediated student interactions and positioning occur in collaborative discussions that utilize participation structures. Our goal, then, is to explore the relationship between these processes, and how they explain the development of positions of authority.

The primary data source utilized here is a 27-minute videotape of the student-led debate. Video was recorded by a mostly fixed Hi-8 camcorder that captured the circularly arranged group from the side and was connected to a radio microphone placed in the center of the group. Video data were coded and organized into an analytical transcript that was divided into columns (Barron & Engle, 2007; Jordan & Henderson, 1995) for words and actions relevant to: (a) making claims about whether orcas are whales or dolphins, and (b) indications of being influenced by particular students; (c) type of evidence used to support those claims, and (d) the perceived authority of sources of evidence; (e) the management of the conversational floor and interactional space; and (e) a miscellaneous column for all other words and actions. Multifunctional words and actions were included in multiple columns. This analytical transcript was originally used for previous analysis that considered multiple factors explaining students' differential influence during this student-led debate (Engle, et al. 2007; Engle, Langer-Osuna, & McKinney de Royston, 2008). Of the four explanatory factors, students' negotiated levels of authority were particularly central. This, in this paper, we go beyond previous analyses to unpack exactly *how* Brian developed a position of authority vis-à-vis other students during this student-led debate.

To understand how Brian developed a position of authority, we coded for events that signaled a shift in each student's position with respect to authority. A person was considered to be positioned with authority on a topic to the degree that he or she was evaluated, acted, or was treated as being credible. A shift in a student's perceived authority could be signaled by statements that contained explicit evaluations of his or her credibility (e.g., "He's features expert, he should know"), statements or actions that evidenced that student being treated as an authority (e.g. one student asking another to give them the correct answer and recording it without request for justification), or the student acting like an authority without anyone contesting it (e.g., "Believe me, I should know" with no one else objecting). For each shift in students' perceived authority, we then noted the student interactions that preceded the shift. Patterns of interaction were compiled into categories (Guba & Lincoln, 1982), which served as a basis for characterizing the interactions that explained the development of Brian's position of authority in this student-led debate.

## Characterizing Changes in Brian's Authority During the Student-Led Debate

We begin with an analytical narrative (Angelillo, Rogoff & Chavajay, 2007) based on the coded, analytical transcript. The narrative describes the key events of the debate relevant to characterizing how Brian's authority developed. Because another student, Samantha, was the most vocal opponent to Brian, changes in her authority, which increasingly became oppositionally paired with Brian's, are also described.

### Early in the Debate, Both Brian and Samantha Positioned with Similar Levels of Authority

At the start of the debate, both Brian and Samantha were generally oriented to as having higher authority than other students in the group. Brian was recognized as the expert on whale anatomical features, his assigned role for the project. In contrast, Samantha was often oriented to by the teacher and other students as a good student who knew things, which was evidenced by another student's comment at one point in the debate that "we all goin' against Samantha; we better be right." (line 420). Samantha's authority also may have been strengthened given her recent influence in the hallway debate about whether to include orcas on their bulletin board. Early in the debate (through line 97), Brian's authority was given additional boosts by the teacher who represented him as having "serious points" that should be considered.

### Brian's Authority Was Challenged

During the next phase of the debate (lines 97-180), Brian's authority was challenged on three occasions when the quality of his evidence was evaluated as being low. First, Brian's and Samantha's credibility were both challenged when the teacher declared that only one student, Jonah, had thus far presented evidence. His arguments had used documentary sources (lines 50-52) while theirs had not. Brian and Samantha then argued back and forth for opposing positions with no resolution while each implied that the other had lower credibility. This started to put their positions of authority in opposition. Next, Brian's authority received an additional downgrade by the teacher who required him to provide a more well-justified argument before she left the group to attend to other groups. As the teacher left, a next student, Toscan, took up Brian's role as features expert by asking him for key facts about whales' anatomy. However, Brian's role as expert was weakened further as Jonah used new documentary evidence to contradict Brian's claims, which Toscan subsequently acknowledged.

### Brian's Authority Steadily Increased

A transition occurred during the next episode (lines 181-379) when Samantha began defending the authority of the trainers. Brian then began arguing for his superior authority on the topic due to his role as features expert. Toscan defended Brian as having greater authority than the trainers, and presumably anyone else who might wish to argue against Brian. Several other students joined Brian's side of the argument during this time. Later, after having difficulties getting selected for turns, Samantha interrupted to argue against Brian's presumptions that he was a scientist and that the trainers were not, but was stopped first by Brian and then Toscan.

In the debate's climax (lines 380-435), Brian made a point against the trainers' authority that was oriented to as high quality and in response three more students publicly declared allegiance to Brian's side while one strengthened her commitment to it, with Brian and his allies celebrating after each success. It began with a rolling chorus of student agreements to Brian's point about the trainers not being experts as one of them confused how many teeth orcas have as compared to dolphins. At this point in the debate, most students had aligned with Brian. Jonah then got the floor and argued that the book he was using as documentary evidence to challenge Brian's credibility offered contradicting information, discrediting the book's authority. Gaining Jonah as an ally was interpreted by Brian and Toscan as meaning they had "won" the argument. Finally, Brian got Jonah to call on a still unaffiliated student, Sione, who also declared allegiance with Brian's side. The teacher then called to the group to wrap up the debate, and Brian concluded, "I rest my case."

## Analyzing the Development of Student Authority

Our analyses, described below, show that Brian became positioned with and sustained relatively greater authority primarily because: (1) he utilized his role as "features expert" to mediate his position of authority vis-à-vis others; and (2) he made effective use of potential allies who took up his position of authority.

## The Role of Expert as a Resource that Brian Used to Position Himself with Authority

Brian engaged in particular forms of interactions that served to maintain his position as expert on whale features and become positioned with relatively greater authority than others: (a) he highlighted the relevance of his role, while (b) personalizing his expertise beyond the role itself; and (c) he simultaneously redefined the trainers as non-scientists, and then (d) linked the discredited trainers with Samantha, which reduced her credibility.

### The Relevance of Brian's Role as Features Expert Was Made Central to the Debate

Brian's role as features expert was germane to the debate that erupted among students after the Marine World fieldtrip. That is, to decide whether orcas were actually whales, an understanding of the anatomical features of whales would be critical. However, other sources of information contradicted his conclusion that, because of similarities of anatomical features, orcas were indeed whales. Students drew on these sources of evidence, which included a book on whales, to argue against Brian's conclusions. Because classroom norms valued backing up arguments with sources of evidence, the information offered in the book became more central to the debate. Indeed, early in the debate when the teacher was still present, she stated that only one student, who had drawn on information from a book, had thus far used any evidence despite the fact that Brian had cited particular features of orcas that were similar to whales. In doing so, Brian's "knowledge" of anatomical features became less central than the information gleaned from the book or even from the trainers during the field trip.

Thus, in order for Brian to claim a position of authority, he needed to make his role as features expert just as central a source of information as books or the trainers. In the excerpt below, Brian positioned his role as features expert as a central *source of information*:

51 Brian: Oh . . it's . . it's kinda it seems like it's not true [*that orcas are dolphins*]
52 because . . um in all the research
53 everything I've done on/ on killer whales . . . .
54 it points to that they are whales
*(+ signifies Brian's gestures for emphasis)*
55 the +blubber . . the+blowhole . . the +eyes . . the +teeth
56 it points out . . that they're a whale

Brian's phrases "in *all* the research" and "*everything* I've done" (lines 52-53) emphasize that he had done a lot of work on his route to becoming an expert and had gathered much information, such that his conclusions (lines 54, 56) ought to be considered credible.

However, his conclusions were challenged by some of the other students who drew on other sources of evidence that suggested that these similarities also extended to dolphins. Thus, as in the following statement, and in many other moments (included in other sections below), Brian positioned his role as features expert itself as making him a central *authority* in the debate:

199 Brian: I study FEATURES . . believe me I should know . . I should know . .

When other students, in particular Samantha, asserted that the trainers were also in positions of authority on the topic by virtue of their profession, Brian sought to expand his expertise by personalizing it beyond the classroom role to Brian *himself* being an authority on the issues relevant to the debate.

## Brian personalized his expertise beyond the role itself

Brian personalized his expertise so it transcended his classroom role as features expert by claiming that he as a person was an authority on issues pertinent to the debate:

226-7 Brian: I've known a lot about whales . before I started researching . .
…
498-9 I:: . know .a lot about whales personally before I even started researching this . .
500 so nothing's that new to me . . . .

Brian defended his position of authority by claiming that his expertise went beyond the role of features expert to include the contents of his individual mind (line 290), which existed prior to the assigned role (lines 498-500), and was inaccessible for critique (line 291-2).

## Brian redefined the trainers' role to that of non-scientist

At the same time, Brian simultaneously defended his position of authority by *redefining the trainers' role* to that of a non-scientist. Brian's role as classroom expert on the anatomical features of whales gave him particular tasks to complete, which included researching, documenting, and sharing information about whale features with his group, the class as a whole, the rest of the school, and any adults who might engage with the class (see Engle & Conant, 2002). Brian's specific tasks were well-defined and easily defended to the other students as engaging in "real research." Brian thus argued that the role of Marine World trainers included little more than performing with dolphins and included little of the real work of scientists like himself.

In the excerpts below, Brian contrasted his and the trainers' roles, asserting that they were not comparable in expertise:

221-3 Brian: I study FEATURES . . believe me I should know . . I should know . .
224-5 they TRAI:N whales . . there's a BIG old difference . .
…
369-70 Brian: okay I'm gonna change and be one of the persons who works at Marine World
371-3 for the whale training . . [ I apply for a job . um . and then my boss tells me
374-6 . . take this stick . . tap [ the water . . and then the whale will come up
…
387-8 and I know a LOT about whales, AND I've REsearched it . .
…
478-9 Brian: like they're really sci:entists /they just apply for a job one day
480 got told a little bit about it .
481 and they've been working with the whales ever since . .

In lines 221-225, Brian directly contrasted the two roles and asserted that he, unlike the trainers, actually studied the anatomical features of whales. The trainers, instead, were simply employees who were told by their employers what to do to make the whale come out of the water during the show, and perhaps told also a "tiny bit" about whales (lines 224, 379-381, 480). Toward the end of the debate, Brian directly challenged their expertise by sarcastically stating, "like they're really scientists" (line 478).

By engaging in such interactions with the group, Brian positioned himself as a more valid source of information than the trainers by virtue of their contrasting roles. Brian's role as features expert positioned him as a "real scientist", while the trainers' role was redefined to include little more than following directions

### The discredited trainers became linked to Samantha's credibility

In the course of the student-led debate, Samantha became the most vocal supporter of the trainers' credibility. She was considered to be someone who gave evidence-backed arguments. Indeed, prior analysis of this data (see Engle, et al., 2007, Appendix A) indicated that Samantha's arguments during the focal student-led debate were more objectively meritorious than Brian's arguments. Thus, in order to reduce Samantha's credibility as a student with valid arguments to consider in the debate, Brian (and his allies) effectively linked Samantha's credibility with that of the discredited trainers. Because the trainers' authority was, in essence, directly in opposition to Brian's, Samantha's authority became oppositionally linked to Brian's.

In the following excerpt, Samantha was positioned as knowing less than Brian, even though at the start of debate, and throughout the year, Samantha was considered to be a student with high academic status.

| | | |
|---|---|---|
| 410-1 | Samantha: | I see (your point of view) but still . . is . that [*the trainers know better than us*] |
| 412-3 | Jonah: | still what? [*Toscan puts hand down, Jonah's is still up*] |
| 414-5 | Brian: | *(to Samantha)* At first you're TOtally convinced that a killer whale is a whale . . . |
| 416-7 | | THEN you find out just from ONE person . from ONE moment . |
| 418 | | that it's a dol]phin. |
| … | | |
| 447-8 | Brian: | [*the trainers know better than*] you do . . better than . YOU do . . |
| 449 | | I::: know what I:'m talkin' about |

The excerpt above followed after Samantha asserted that the students should believe the trainers because they likely knew more about whales and dolphins than the students did. Some of the students then pointed out mistakes they recalled the trainers making during their performance in order to discredit them. Samantha conceded their point and had begun formulating a counter-response (lines 411-412) but was then interrupted and criticized for believing, not an expert, but rather "one person from one moment" (lines 416-417). Brian then shifted the nature of the criticism to be about Samantha herself, stating that the trainers, if anything, knew more than she did (lines 447-448) but not more than he did (line 449). In doing so, Brian first linked Samantha's and the trainers' credibility, and then demoted Samantha as knowing less than even the discredited trainers.

## Allies Took Up and Supported Brian's Position of Authority

Brian's moves to position himself with greater authority than even adult professionals could not be successful by themselves. Other students could have rejected his attempts to position himself in such ways, negotiated an alternative positioning with him, or taken up his proposed positioning as having greater authority (Clark, 1996).

In order to encourage the latter to occur, Brian engaged in several kinds of interactions that served to gain and maintain allies, including taking credit for arguments stated by others, and rousing students' desires to be his ally rather than foe. We then see evidence that students rallied in support of Brian's credibility, joining in the criticisms of the trainers, Samantha, and even the book author, while aligning with and building on Brian's arguments or the arguments of others that Brian had claimed as his.

### Brian gained authority by taking credit for others' points that supported his side of the debate

Brian often positioned himself with authority during the debate, which was at times taken up or challenged by the other students. Brian was able to increase uptake of his position of authority by aligning with, and taking credit for, points made by others that had been oriented to as compelling in the debate. In the following excerpt, Brian claimed another student's point and was later credited with that point by a third student:

| | | |
|---|---|---|
| 213-4 | Jonelle: | if . if they're dolphins why do they call them killer whales? . . |
| 215 | | why don't they call them [ (killer dolphins?) |
| 216 | Brian: | My point exactly . . |
| … | | |
| 234 | Toscan: | ok, Brian has a very good point . . . . |
| 235-6 | | when . . the killer whales, if they're called . |
| 237-8 | | and Jonelle, if the killer whales are called killer whales . |
| 239 | | why are they saying NOW that they're dolphins |

Jonelle argued that orcas, also known as killer whales, were likely to be whales because their name includes the term "whales," not "dolphins" (lines 213-215). Brian immediately then exclaimed, "my point exactly" (line 216) claiming the idea as his. Moments later, Toscan attributed the point to Brian (lines 234) and repeated the argument, mentioning Jonelle as linked to this point only in passing (line 237).

During the debate Brian frequently used the phrase "my point exactly" or later simply "exactly" (lines 41, 44, 216, 620-621, 659, 695, 721) in response to others' points for the orcas-as-whales view and then got the last word at the very end of the debate, saying, "I rest my case" (line 836). In essence, Brian successfully took ownership of many of the arguments offered by others that supported his side of the debate, positioning himself as the leader of the "orcas as whales" side, and the other students as merely supporting *his side* of the debate.

### Brian created in students a desire to align with him rather than against him

In addition to gaining allies by essentially becoming the figurehead of successful points, Brian also created in students a desire to align with him by dramatically celebrating when students took up his position through cheers and high-fives while also creating an intimidating climate for disagreeing.

Brian loudly and excitedly celebrated when students announced either switching to or confirming support for his side of the student-led debate. For instance, in response to a claim that supported Brian's attempts to discredit the trainers, Brian loudly and excitedly exclaimed, "Yup, yeah you're right! You're right! You're hecka right!" Moments later, Brian exclaimed, "whooo!" in support of another student contributing arguments that also positioned her on the orcas-as-whales side. Directly after this, several students shot their hands up to speak next. From that point on, the majority of students in the group shared cheers, high-fives, and animated talk as they contributed to the orcas-as-whales side of the argument, creating a snowball effect of support for Brian's position of authority in the debate (Anderson, et al., 2001).

For instance, in the excerpt below, Brian dramatically celebrated after Jonah read information from the book *Album of Whales* that could be understood to support the claim that orcas are not dolphins:

647-8 Jonah: *(holding up book)* it says right here . that um . wha/ that . um
649-50 that wha/ . that um dolphins are beaked wha/ . beaked whales . . a family
651-2 *(looks up from book)* killer whales are not even/ . don't even have have a beak
653 Brian: A::::::::::::::::::::H *(screeches; Brian jumps up and down, in and out of chair)*
654 *(Brian high five's Toscan and Sione)* . . . . We WON!
…
665-7 . . we convinced him . we convinced him we convinced him
…
672 *(Brian jumps up from chair, turns around dancing in place three times, lifting right and left arms in air alternately, and says in sing-song voice)*
673 HALLELUIA, that's the way

Jonah offered evidence that dolphins are considered "beaked whales" (lines 647-650) and noted that orcas do not have a beak (line 651-652). This was Jonah's first contribution supporting the orcas-as-whales side; his previous contributions, which all used high-status and high-quality documentary evidence, tended to support the opposing view. Brian dramatically celebrated Jonah's contribution, which he interpreted as signaling that Jonah had switched sides. Brian jumped in and out of his chair, screamed, gave high-fives to nearby students, danced around, and even yelled out, "Halleluia" in praise of gaining a new ally in Jonah, also a high-status student.

These celebrations were coupled with harsh responses to Samantha's challenges of Brian's position of authority, which may have dissuaded other students from also challenging him. For instance, Brian often groaned in protest, or aggressively pointed or gestured to Samantha when she challenged his credibility. Over time, other students, as examined below, even jumped on the bandwagon of trying to intimidate Samantha.

### Allies tag-team in a snowballing increase of support for Brian's credibility and against others

Most of the students ultimately supported Brian's position of authority in the debate. They positioned his role as features expert as central, discredited the trainers and the book author, and took up his arguments.

For instance, in the excerpt below, Toscan defended the centrality of Brian's role as features expert:

253 Toscan: if Brian is studying features . .
254 why didn't HE get the information before Jonah did . .
255 because he was studying it a lot more and . .
256 every book that he got said whales and now they're saying . .
257 "OH yeah they're dolphins" . .

Toscan challenged the credibility of Jonah's argument, based on evidence he found in a book, by drawing on Brian's role as features expert. He claimed that Brian, as features expert (line 218) had researched far more than Jonah (or other students) had (line 220), and had considered more than just one book (line 221). In doing so, he took up Brian's role as features expert as central to the argument, and Brian as a credible source of information.

Students also joined Brian's attempts to discredit the trainers. In the excerpt below, Liana built on Brian's assertion that the trainers were not really experts and therefore not believable:

607 Liana: [OH! . . oh . . oh . . *(raising her hand excitedly)*
608 A:::ND . she probably didn't even know .
609-10 because she was probably just there, for a little whi::le .

612 and then she just made it [up

Liana contributed to Brian's effort to re-define the role of the trainers as non-scientists. She claimed that the trainer "probably didn't even know", had been there for just "a little while", and had "made up" the information offered at the show (lines 608-610). Finally, the other students joined Brian in trying (though unsuccessfully) to intimidate Samantha to change her position. In this excerpt from late in the debate, Samantha was hassled in multiple ways for continuing to disagree with Brian:

425 Samantha: okay . so . well . I still don't believe it
426-7 becu:z . I don't think she said that, "I think" . . . ]
428 Jonah: [she did too]
429 Samantha: [what I remember is .
430-1 her saying that . . that "we have figured out that . . . . . .
432-3 Brian: *(holding tape recorder to Samantha)* Talk! . . . . talk . . . . ]
434 Sione: *(holding microphone within inches of Samantha's face)* [Talk! . .
435-6 Brian: *(to Samantha)* C'mon, just forget it.
437 Samantha: the whale/ ok . the killer whale
441 is a probably . a dolphin becuz .
442 I think they know better what they're talking about than we do/]
443 Jonah: [PRO:BABLY?
444-5 Brian: NO! . . (MAKE THAT BETTER THAN . YO::U! DO! . . ]
446 Jonah: [*chuckling*] probably

Samantha was frequently interrupted here (lines 428, 432-436, 443-445) and earlier, but she successfully reclaimed the floor (lines 419-425, 429-431, 437-442). In response, Brian and his allies sought to intimidate Samantha by placing the researchers' microphone directly in her face, spatially crowding her, and yelling at her to "talk!" (lines 432-434, and later), while Brian then calmly and almost dismissively urged her to "c'mon, just forget it" (line 435-436). When Samantha persevered (lines 437-442), Jonah and Brian ridiculed her.

As dramatic as these interactions may have been, once the students were regrouped as a class with the teacher, opportunities to address Brian's domination arose. In the ensuing weeks of the unit as the students continued to debate the question, the teacher continually emphasized the importance of considering evidence and discouraged students, including Brian, from declaring victory. Students found additional evidence for Samantha's side. By the end of the unit when Brian wrote about the controversy in his part of the group's report, he represented it as a stalemate with evidence for both sides. This made sense given that orcas can be considered to be both whales and dolphins (orcas are in the "dolphin" [delphinidae] family, which is itself within the sub-order glossed as being the "toothed whales" [odontocetes]). Thus Brian's strong position of authority, perceived as even greater than the authority of the trainers and the book author within this one debate, did not prevent these students from productively engaging in the issue over the long haul (see Engle & Conant, 2002 for more)

## Discussion

Participation structures, such as student roles that make use of shared expertise, not only mediate students' engagement in collaborative work, but also their positions of authority in relation to other students and beyond. In the case of Brian, he was positioned with authority on whale features because of his assigned role, but it went beyond simply this to afford a positional identity as not only a central and credible student, but also as a "real" expert on par with adult professionals (Langer-Osuna, 2009; Wortham, 2004).

This paper explored how particular participation structures, such as distributed expertise (Brown et al. 1993) can support the development of student authority, and identities of power. These identities may be critical to support valued learning processes, such as students becoming authors of ideas, debating and coming to consensus on student-authored ideas, and using sources of evidence from adult professionals to think through the reasonableness of ideas without passively aligning to the authority of such sources (Inglis & Mejia-Ramos, 2009). However, student identities as authorities can also lead to bad outcomes when students do not sufficiently hold themselves accountable to others and disciplinary norms (Engle & Conant, 2002), resulting in the propagation of invalid ideas and reasoning (Brown & Campione, 1994).

But participation structures are not the only factor to consider with respect to how student authority becomes constructed in classroom interaction. Engle, Langer-Osuna, and McKinney de Royston (2008) offer a model that depicts the dynamic interactions of four negotiated factors that affect how students gain influence over others, based in part on the student-led debate discussed here. These factors include students' perceived authority, but also the perceived merit of their arguments as well as their access to the conversational floor and interactional space. For instance, when Brian's authority was challenged early in the discussion, he was less frequently chosen to speak. However, Brian then worked to increase his access to interactional space through dramatic gesturing, and otherwise spatially gaining attention. In doing so, he gained access to the discussion, which he then used to strengthen his position of authority in the ways we have analyzed here. We believe that such dynamics within and between authority and other factors will also be relevant to other student discussions .

## References

Anderson, R.C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., Tillmanns, M., & Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction, 19*(1), 1-46.

Angelillo, C., Rogoff, B. & Chavajay, P. (2007) Examining shared endeavors by abstracting video coding schemes with fidelity to cases. In R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences*.

Barron, B. & Engle, R. A. (2007). Analyzing data derived from video records. In S. J. Derry (Ed.), *Guidelines for video research in education: Recommendations from an expert panel* (pp. 24-43, 79-80). Chicago: Data Research Development Center. Available at: http://drdc.uchicago.edu/what/video-research.html.

Bianchini, J.A. (1999). From here to equity: The influence of status on students access to and understanding of science. *Science Education, 83,* 577-601.

Brown, A.L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J.C. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.) *Distributed cognitions: Psychological and educational considerations.*

Brown, A.L., & Campione, J.C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229–270). Cambridge, MA: MIT Press/Bradford Books.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cohen, E. G. & Lotan, R. A. (1995). Producing equal-status interaction in the heterogeneous classroom. *American Educational Research Journal, 32*(1), 99-120.

Engle, R. A. & Conant, F. C. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20(4),* 399-483.

Engle, R. A., Langer-Osuna, J. & McKinney de Royston, M. (2008). Toward a model of differential influence in discussions: Negotiating quality, authority, and access within a heated classroom argument. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2010-2015). Austin, TX: Cognitive Science Society.

Engle, R. A., McKinney de Royston, M., Langer-Osuna, J., Bergan, J. & Mazzei, P. (2007, April). *From positioning to differential influence in a student-led argument: Students negotiating authority, the conversational floor, and interactional space*. Paper presented at AERA, Chicago.

Esmonde, I. & Langer-Osuna, J. (2009). *Power plays: How students in a small group negotiated mathematical discussions in hybrid spaces*. Manuscript under review.

Gresalfi, M.S., Martin, T., Hand, V., Greeno, J. (2009). Constructing competence: An analysis of student participation in the activity systems of mathematics. *Educational Studies in Mathematics, 70*(1), 49-70

Guba, E.G. & Lincoln, Y.S.(1982). Epistemological and methodological bases of naturalistic inquiry. *Educational communication and technology, 30*(4), 233

Herrenkohl, L. R. & Guerra, M. R. (1998). Participant structures, scientific discourse, and student engagement in fourth grade. *Cognition and Instruction, 16*(4), 431-473.

Inglis, M. & Mejia-Ramos, J.P. (2009). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction, 27*(1), 25-50.

Jordan, B. & Henderson, A. (1995). Interaction Analysis: Foundations and practice. *The Journal of the Learning Sciences (4)*1, 39-103.

Kurth, L.A., Anderson, C.W., & Palinscar, A.S. (2002). The case of Carla: Dilemmas of helping *all* students to understand science. *Science Education, 86:* 287-313.

Langer-Osuna, J.M. (2009). *The algebra of identity: Toward a new framework for understanding the co-construction of identity and engagement in mathematics classrooms*. Unpublished doctoral dissertation, University of California, Berkeley.

Langer-Osuna, J. (2007). Toward a framework for the co-construction of learning and identity in mathematics classrooms. *Online papers of the Second Annual Socio-cultural Theory in Educational Research and Practice Conference: Theory, Identity and Learning*, Manchester, England, available at http://www.education.manchester.ac.uk/research/centres/lta/LTAResearch/SocioculturalTheoryInterestGroupScTiG/SocioculturalTheoryinEducationConference2007/Conferencepapers/GroupFivePapers/

Saxe, G. B., Shaughnessy, M. M., Shannon, A., Langer-Osuna, J.M., Chinn, R., & Gearhart, M. (2007). Learning about fractions as points on a number Line. In W. G. Martin, M. E. Strutchens, & P. C. Elliott (Eds.), *The learning of mathematics: Sixty-ninth yearbook* (pp. 221-237). Reston, VA: National Council of Teachers of Mathematics

Wortham, S. (2004). From good student to outcast: The emergence of classroom identity. *Ethos, 32*(2), 164-187.

Yin, R.K. (2003). *Case study research: Design and methods* (3rd ed., Vol. 5). Sage Publications.

# Digital Video Tools in the Classroom: Empirical Studies on Constructivist Learning with Audio-visual Media in the Domain of History

Carmen Zahn, Karsten Krauskopf, Friedrich W. Hesse, Knowledge Media Research Center, Tuebingen, Germany
c.zahn@iwm-kmrc.de, k.krauskopf@iwm-kmrc.de, f.hesse@iwm-kmrc.de
Roy Pea, Stanford Center for Innovations in Learning (SCIL), Stanford, CA, USA, roypea@stanford.edu

**Abstract:** This paper presents empirical evidence concerning the application of digital video technologies for creating design-based learning environments for middle school students. In three studies we show how the affordances and constraints of digital video technology can support for students their (a) cognitive (b) action related and (c) socio-cognitive learning processes in the domain of history. We present both quantitative and qualitative data. We also present initial evidence for the role of complementary support by explicit instructional guidance. Results are discussed with regard to practical implications.

## Introduction

Audiovisual media provide an important resource for classroom learning. Yet, films and videos are often used in 'suboptimal' ways (Hobbs, 2006). For example, they are shown to classes in a passive, TV-like manner, without clear-cut educational goals and without fostering students' learning activities, dialogue, knowledge construction, or critical thinking. From a psychological perspective, this kind of usage limits the effectiveness of audiovisual materials for learning, and promotes instead a tendency towards superficial cognitive processing (Salomon, 1984) and oversimplification (Spiro, Collins, & Ramchandran, 2007; Spiro, Feltovich, Jacobson, & Coulson, 1991). While this might be only unfortunate in some cases, it can actually be problematic in others, such as in the complex domain of history (Spiro, et al., 2007). In history learning, media sources including historical films, newsreels, TV-products and videos can help to reconstruct historical processes and events (e.g., Wineburg, 1991). Yet, the sources themselves - rather than merely presenting "facts" - are themselves parts and results of the historical processes from which they have originated, and as such they reflect certain perspectives (sometimes subtley as in news shows or educational films, sometimes obviously or even deliberately as in propaganda films). Hence, historical films and videos need to be understood by students exactly in this way: as historical "constructs" (Wineburg, 1991).

How can such an understanding be achieved? Educators have suggested domain-specific strategies, such as "de-construction and re-construction" of historical films (Krammer, 2006). Taking a general theoretical stance in his important work on cognitive flexibility, Spiro has suggested using hypermedia technology as a supportive tool for multithematic exploration and processes of active and constructive work with video in ill-structured domains. Likewise, in a similar cognitive-constructivist framework for the use of video in education (Goldman, Pea, Barron, & Derry, 2007), video analysis activities with advanced digital video tools have been acknowledged as supporting "perspectivity" (Goldman, 2004), inquiry, and specific (socio-)cognitive processes of students (Smith & Blankenship, 2001). This advanced category of digital video tools (some of them originally created for professional research purposes) refers to software applications which provide a variety of possibilities for editing, contextualizing or analyzing digital videos (cropping, annotating, commenting, tagging, or integrating hyperlinks). Such functions could be restructured for youthful learners in classrooms so that they could either produce their own videos or remix video contents (as, for example, from YouTube).

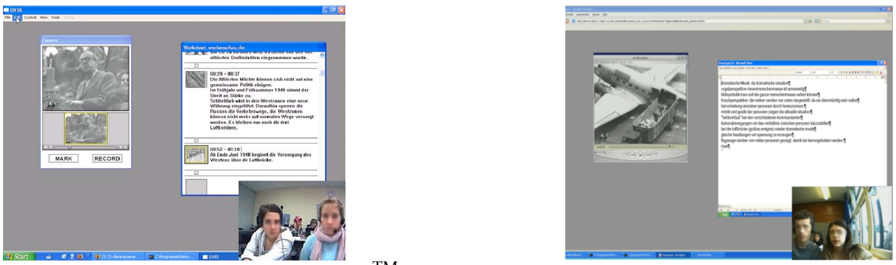


Figure 1. Screenshots of a DIVER™ worksheet (left) and the video player plus text editor condition (right) including video feed of dyads' interactions (bottom right).

In this paper, we present empirical research on the meaningful integration of digital video tools into history education in secondary school. We conducted a series of three empirical studies with 11th grade students in Germany. Our research rests on two assumptions: (1) the assumption that the affordances and constraints of digital video technologies—particularly constructive video tools allowing for annotating, editing (e.g., zooming and cropping), and re-sequencing—have opened up new vistas for making video accessible to constructivist learning in school-based education and beyond (Zahn et al., 2005) and (2) the general assumption that framing such video tools in constructivist design tasks can be effective (Zahn, Krauskopf, Pea & Hesse, in press). We worked with a video tool which has been created specifically to support both cognitive processes and collaborative learning with digital videos (the DIVER/WebDIVER™ system, Pea et al., 2004, see Figure 1). It supports cognitive processes such as focusing attention within a complex and dynamic visual array. It includes a set of specific editing capabilities which enable users to extract, rearrange and comment on scenes or video segments aiming at the support of important visual analytic skills of learners, such as focused observation. The selected and commented items are called *dives*, based on users "diving" into a video by zooming in on scenes and details of the video by controlling a selection frame or "virtual viewfinder". A dive consists of re-orderable "panels", i.e., clips and corresponding comments that are shown in a separate column next to the source video. As a result, sequences and details of a video are not only being cognitively selected, but these video selections and interpretive annotations become new, replayable artifacts that have persistent references as URLs. This is why the system is also considered a *collaborative* tool: These replayable artifacts can be collaboratively produced and shared with others for joint reflection, analysis and discussion ("Guided Noticing"™). In transferring these design rationales underlying the DIVER/WebDIVER™ system or similar tools to the classroom (Zahn et al., 2005), we investigate in our research both the new opportunities and the new challenges arising for students, and ultimately teachers. From our point of view , there are many opportunities to create better conditions for active learning with video materials and relate them to effective processes of knowledge building (Scardamalia & Bereiter, 2006).For instance, students may work creatively with digital videos, which are already available as audio-visual sources online. They can suggest new ways to approach these media and thus enable subsequent users to integrate the reflections of others, which ultimately leads to the development of core competences, such as historical reasoning, advanced expertise (Scardamalia & Bereiter, 2006) or new media literacies (Jenkins, Clinton, Weigel, & Robison, 2006). The main challenge is to frame these collaborative tools with the right tasks to create powerful learning environments, in order to reach the learning goals during classroom- based instruction.

## Visual Design to Frame Learning with Digital Video Tools

To investigate both the opportunities and the challenges of using advanced video tools for constructivist learning in history lessons, we apply design approaches to learning, i.e., concepts of learning by creating artifacts (e.g., learning through design, Kafai & Resnick, 1996; learning by design, Kolodner et al., 2003). We consider the design approaches especially suitable in our case, because they call for activity- and project-oriented learning with media. In this process, students can autonomously create something that seems authentic, meaningful and consequential for them, which gives them the opportunity to experience themselves as competent and intrinsically motivated learners (Beichner, 1994). In addition, from a cognitive perspective, a deeper understanding of the complex topic area and video content can be expected: Acting as designers, students are creating a new video-based information structure. This requires them to select, compare and reflect on the content they wish to present and on how to present it, which results in a positive influence on knowledge acquisition for a complex topic, as in writing (Hayes, 1996). Based on the work of Lehrer and Erickson (Erickson & Lehrer, 1998; Lehrer, Erickson, & Connell, 1994), the ideal structure of a design process includes planning, transformation, evaluation and revision. Design projects thus should support the development of general problem solving skills as well as specific skills related to media competence. Learners not only need to plan the information structure they are going to create, they also must creatively integrate different media formats in order to combine them effectively for a sensible form of the information presented. In an attempt to integrate the cognitive models of writing and design with constructionist perspectives, and in order to apply them to the case of video-based design, we proposed elsewhere a tentative model of collaborative visual design as a process of joint dual space problem solving (Zahn, Krauskopf, Hesse, & Pea, 2009). In this model, we hypothesize visual design to be a collaborative problem solving process involving intensive interactions between video content and form (audience-related goals), in a rhetorical problem space. Based on this model, we shaped our research around the following research questions:

- How do the affordances of advanced digital video tools impact (a) cognitive (b) action related and (c) socio-cognitive processes of students learning a historical topic?
- How do students approach visual design tasks in a real history class?
- When and how can explicit instructional processes optimize the implementation and utilization of digital video technologies?

For the purpose of our empirical research, we developed a prototypic visual design task for German history lessons, where students were asked to rework a video source showing an historical newsreel on the 1948 Berlin blockade. With varying digital video tools at hand, learners elaborated on the video source by integrating additional information provided, with the goal of creating an artifact that could be published online for future (peer-)learners. The task included critical reflection and comments on the historical content as well as on newsreels as a means of propaganda in former times, based on examples from the source video. Whether the affordances of digital video tools used in the context of our creative task would clearly have a positive effect on student understanding in a real history class was an open empirical question. It was also an open question whether the affordances and constraints of different digital video tools differentially influence student learning and if so, whether they would affect specific task elements. In general, we assumed the visual design task and the affordances of digital video tools would have an impact on three levels: (a) the cognitive impact was assessed by multiple choice items tapping historical knowledge and transfer tasks for critical film analysis. We investigated the design products of participants as indicators of (b) action-related outcomes by considering their number of selected scenes and sequences, precision (length of sequence and size of detail), and structuring (deviations from the source video's chronology). Finally, we analyzed the dyads' taped interactions as indicators of (c) socio-cognitive effects by using a coding scheme that captures students' problem solving behaviors (planning, executing, evaluation, revision), collaborative behaviors, and content discussed (the original newsreel-video, the historical events, or their design task). In addition, we analyzed selected case examples of dyads' interactions to provide empirical evidence for the impact of the affordances of the advanced digital tools on collaboration.

The visual design task was tested in a lab-experiment (see Zahn, Pea, Hesse, & Rosen, in press) under controlled conditions to investigate its general effectiveness. Also, the specific effects of the affordances of the digital video tool WebDIVER™, in contrast to a control condition using "simpler" technology (video-player & text-editor) had to be proven before we took the procedure into studies with high school students in the classroom. The sample of this first study consisted of 24 dyads of psychology students (mean age = 22.2, *SD*= 4.8, 68.8% female). Generally the results revealed high appraisal of the task and significant positive mediating effects of the WebDIVER™ video tool on all three levels of outcomes described above: Dyads supported by WebDIVER™ showed higher factual knowledge and film analysis skills, design products of higher quality, and more communication time devoted to task relevant content (historical context and design decisions). In sum, the results showed that the influence of the video tools extended to the learners' socio-cognitive processes and focused their interactions on the task. Qualitative case analysis provided evidence for direct effects of the video tools on conversations, i.e., tool affordances implicitly guided learners' elaboration on the source video.

## Field Study in Class

Following up on the results from this preliminary laboratory study, we took the experimental design into German secondary classrooms. The experimental sessions were conducted in schools spanning over two subsequent 45 minute units using a portable "notebook classroom" we provided for the study. The students were alerted by teachers that our task would be integrated into the current history curriculum. In this study we manipulated two experimental between-subject factors in a two-way factorial design. The first factor (technology) was the digital video tool provided to the students (digital video tool DIVERTM vs. simpler software solution, i.e. video player combined with text-editor). The first factor was chosen to test the ecological validity of the results found in the prior lab experiment on the specific effects of the affordances of the digital video tool WebDIVER™. By investigating effects of tool affordances, we wanted to address the important issue of whether or not an editing tool originally designed for researchers and educators, when reapplied as a tool for learning in the classroom, would be beneficial for students in comparison to a common (and maybe more familiar) technology. The second factor (metaphor) was the appeal to students' media-related rhetorical concepts –their rhetorical concepts were either in line with or different from the system architecture of the digital video tool). By introducing this second factor, we wanted to tap into the question of explicit instructional support of the complex design task as opposed to the implicit guidance by the digital tools. We based this factor of metaphor on the work of Stahl and Bromme (2004), who investigated students designing hypertexts and reported that, due to the novelty of the medium hypertext, student designers cannot be assumed to have firmly established media-related rhetorical concepts which form a design goal. Therefore, in the first condition we prompted students with a metaphor of "video dives" epitomizing the concept of a medium for selecting ("diving into") visual information. In the second condition we prompted students with a metaphor of "annotated movies" epitomizing the concept of a medium for integrating information elements.

Altogether 234 students from 8 classes in 4 German secondary schools participated. Due to technical problems, data of 24 participants (12 dyads) could not be saved and were not included in the analyses. In the end, data from 111 dyads of students were analyzed (mean age =15.9, *SD*=0.78, 60.9% female). We are aware that our data has a multilevel structure; however, due to the sample size and small number of classes, we based our analyses on the dyadic level. Gender composition of the dyads was equally distributed over the four conditions. Furthermore, experimental groups did not differ with regard to their experiences with digital media, personal computers, or prior knowledge of post-war German history. However, the groups using the DIVER™ tool reported higher initial interest in post-war German history. Thus, all reported analyses were additionally run with interest as covariate without changing the results.

In general, the design task was rated by participants as being interesting in general and proved applicable in the classroom. Also, with regard to cognitive outcomes, all students improved in their factual knowledge of the historical content, measured by a multiple-choice test before ($M$= 45.20% correct, $SD$ =15.50) and after the design task ($M$= 65.00% correct, $SD$=8.20), $F(1, 107)=230.83$, $p<.05$, partial $\eta^2= .68$. However, there were no differential effects of the independent variables. The manipulation tapping students' media concepts showed no effects on either outcome variable, nor were there any significant interactions ($F<1$).

With regard to the action-related outcomes, students using the advanced digital video technology (DIVER™) created more sophisticated design products. They selected significantly more scenes and more details within these scenes compared to the control group. Furthermore, the DIVER™ tool increased the number of deviations from the chronological order of the source video when students made the selections, which indicates more autonomous design in the experimental compared to the control condition. Altogether, this indicates that the advanced digital video tool fostered a higher quality of the design products.

In line with these findings, dyads using the DIVER™ tool seem to have acted more autonomously during collaboration: In order to analyze the socio-cognitive impact of the affordances of digital video tools on collaborative design, the interactions of a random subsample ($N$=14 dyads) were coded for (a) content of collaborative talk and (b) problem solving behavior. 20% of the videos were coded by a second rater and according to Asendorpf and Wallbott (1979) we computed Cronbach's α for ratio scales (here the aggregated time of students spent on the different activities) as the agreement measure. Overall, high inter-rater reliability was found, all αs > .81.

The proportions of content discussed during the design task, relative to overall collaboration time, show a beneficial influence of the advanced technology. Dyads working with DIVER™ discussed the historical content more substantially ($M$=14.22% of time, $SD$=7.37) and their actual design task ($M$=16.45%, $SD$=5.96), than did dyads in the control condition ($M$=5.33%, $SD$=6.56), $F(1, 107)=6.42$, $p< .05$, partial $\eta^2= .39$, respectively ($M$=7.23%, $SD$=6.55) $F(1, 107)= 7.78$, $p< .05$, partial $\eta^2= .44$. In contrast, students working with the less sophisticated video player and text-editor showed a tendency to discuss the source video itself more ($M$=18.64%, $SD$=8.36) than did students using DIVER™ ($M$=3.49%, $SD$=18.03), $F(1, 107)=3.37$, $p< .10$, partial $\eta^2= .25$. Overall, this indicates that the affordances of the advanced digital video tool influenced the dyads' collaboration on a socio-cognitive level and may have fostered task-relevant discussion about historical content and design.

Students' general problem-solving behavior, however, was influenced neither by the design tools nor the prompted metaphors. All dyads directly engaged in executing the task (80.5% in general) and devoted less than 4% of their time to planning and evaluation. Analyses revealed a tendency for students to request help less often when using DIVER™ (1.63% time working on task, $SD$= 2.13) than in the control condition ($M$=5.32%, $SD$= 4.04), $F(1, 10)= 4.21$, $p<.10$, partial $\eta^2= .26$). We interpret this as an indicator for a stronger implicit guidance of the advanced digital tool in situations when a task is not explained step by step but determined by the goal. The coding of the video interactions also showed that overall "other" activities of learning dyads made up less than 5% of the time on task. These activities included off-task behavior, such as chatting ($M$=3.90%, $SD$=3.58) and problems concerning the technology ($M$=3.67%, $SD$=2.82) and did not differ between the four conditions, all $F$s < 1, ns. Therefore, any differences in students' learning behaviors cannot be attributed to differences in the difficulty of operating the two technologies.

However, especially taking into account the limited time that students could spend on the design task, the quantitative indicators for learning outcomes document effects, *not* processes. In order to provide evidence windows to substantiate our assumptions (see above: advanced digital video tools may support collaboration on a socio-cognitive level), we conducted additional case analyses focusing on giving what Barron (2003) calls a "localized account" on how the dyads integrated technology affordances in their conversations for their design-related interactions. Such evidence was already found in the prior laboratory study (Zahn et al., in press), where we identified processes of "guided noticing" (Pea, 2006), which emerged as interactional patterns in dyads elaborating on a newsreel with WebDIVER™. In the present paper, we provide evidence for similar findings under the conditions of the real "noisy" classroom setting. We compare two exemplary episodes for processes in two dyads working with either DIVER™ (see Table 1) or the simpler technology of the control condition (see Table 2). These dyads performed similarly well on the task and the post-tests. The episodes illustrate how the students from each dyad collaboratively work on the *same* scene in the historical newsreel video, but with *different* orientations: Dyad 1 (Students A and B working with DIVER™) discusses a design problem thereby elaborating on historical content, while dyad 2 (Students C and D working with a video player combined with a text-editor) discusses descriptive features of the source video while also sticking to its chronology.

In dyad 1, during most of the episode the two students skim through the dive panels they had created earlier, which results in several cycles of "guided noticing" (Pea, 2006). The transcript provided in Table 1 illustrates how A guides attention to a previously selected picture. The joint attention of students A and B is also implicitly guided by the technology's affordances: In the beginning of the episode (lines 1-9) the students seek a picture as a focus of their joint attention and select it for their design product (while deleting another prior selection). In the middle of the episode the question of student B about a necessary design decision (where to assign the caption "Air Lift", line 11), initiates a short discussion (lines 12-16), which ultimately leads to meaning-making about what "Air Lift" means (line 12) and what "Candy Bombers" are (lines 14-16). In other words: understanding of the historic event (the aim of supplying the population separate from the planes as a means of reaching this aim and their name "Candy Bombers") results from a simple question (line 11), which resulted from the necessity to write and place a caption. This provides an example of how the tool's affordance impacts socio-cognitive processes, joint meaning-making and the establishment of common ground, finally leading to creating a design product. There is an additional factor suggested by the excerpt presented here: While the dyad diverges from the central goal in this episode—specifying what is meant by the term Air Lift—the dive panels created earlier repeatedly guide joint attention back on this goal. Students A and B finally record the different ideas emerging during their conversation, in other words, they arrive at closure, manifested in their design product as a replayable artifact.

Table 1: Exemplary Episode of Dyad 1 (Working with DIVER™)

| *Timestamp 00:13:27* | | | |
|---|---|---|---|
| **Line** | **Student** | **Utterance** | **Actions and written comments** |
| 1 | A | Where is the picture?<br>There it is. | B scrolls up to the specified dive |
| 2 | B | Yes. … | |
| 3 | A | What's that? | |
| 4 | B | That's … [inaudible] … write "Air Lift" | |
| 5 | A | I would leave that the way it is. | |
| 6 | B | What is that? | |
| 7 | A | Yes, that is a good picture. …<br><br>But you can delete the one below. | A refers to a selected detail of an image showing a group of men carrying bags of supplies |
| 8 | B | Too bad, actually. | B deletes the indicated image |
| 9 | B | Um … (verbalizes while typing) Supplies provided by the Allied Forces? | B selects a dive showing two men unloading a truck with supplies and types "Supplies provided by the Allied Forces" |
| 10 | A | Mmhmm | |
| 11 | B | And where do we write "Air Lift"? | |
| 12 | A | Well, Air Lift means the supply with … carried out by the airplanes. | |
| 13 | B | But it has to be mentioned somewhere. | |
| 14 | A | Well no. … Because everything is about the Air Lift … Nee! … Well, or we add "Candy Bomber" to a picture of a plane. | |
| 15 | B | Were they all named like that? | Screens through the selection |
| 16 | A | Well no, but this is what they were called … well, colloquially. | |
| 17 | B | All of them, really? | |
| | | | A selects a dive of a still showing a group of airplanes on the ground. |
| 18 | A | Yes. … | |
| 19 | A | Ah … I thought you had deleted that. | |
| 20 | B | I did. | |
| 21 | A | Ah, you it is not possible to go down here like this.<br>What is that? … people …<br>That one, we can also delete. | A deletes a dive |
| *Timestamp 15:53 – 16:09* | | *Experimenter announces that everybody should write their final comments and prepare to proceed to the following tasks.* | |

In comparison, as can be seen in Table 2, the conversation of dyad 2 is focused on describing the source video's surface features instead of dealing with design issues and the historical content. The episode starts with C guiding D's attention to the music on the audio track (line 1). After an initial affirmative reaction (line 2), D draws the joint attention away from the music towards the pictures, and states that the airplanes are shown from different camera angles (line 2). C writes down this discovery and continues D's initiation of a deeper interpretation (lines 4-6), which D incorporates in the written comment (line 5). Next, D focuses their discussion on the word choice (lines 10-12), draws the focus away from interpretation back to the observable surface level (line 6), and C corrects his former writing. They then return to the issue of the music (line 8) with which they started out, and they describe the accompanying video images (lines 9-11). Continuing to refer to the music accompanying the final scene, D suggests an interpretation (line 14), yet C objects. In sum, while the two students from dyad 2 initiate meaning-making twice in this episode, they do not succeed in achieving true common ground or finally in recording the ideas that emerged during their conversation. In other words, the dyad does not arrive at agreed-upon meanings to be manifested in their design product. Furthermore, they do not elaborate on content or design, but their discussions remains on a descriptive level with a strong orientation towards the surface features of the original source video. This is in line with the quantitative findings reported above (proportions of content discussed during the design task). Moreover, with regard to the socio-cognitive effects of the control condition technology, we found another aspect. As suggested by this excerpt of this dyad's conversation, the technology affords the students to *proceed according to the chronology* of the source video. This contrasts with the conversation of dyad 1, in which the students skimmed through the source video *following an idea.*

Table 2: Exemplary Episode of Dyad 2 (Working with Video Player combined with Text Editor)

| *Timestamp 00:20:29* | | | |
|---|---|---|---|
| **Line** | **Student** | **Utterance** | **Actions and written comments** |
| 1 | C | Military marches. | C stops the video |
| 2 | D | Mhm [affirmative]<br>Let's write that the planes are being shown from many different angles [incomprehensible] concentrate | C starts to type "planes are shown from many different angles" |
| 3 | D | Now you typed perp. [laughs] | C corrects the typing error perp. to pers. |
| 4 | D | „in order to" ..., come on, say something | |
| 5 | C | „in order to ...<br>make people concentrate on them" | C continues to type "in order to make people" |
| 6 | D | Yes …<br>„in order to highlight them" is better. | C changes the comment into "in order to highlight them." |
| 7 | C | An what do we write here? | |
| 8 | D | „military marches" | C types "military marches" |
| 9 | C | "while …" do you get that? | |
| 10 | D | Yes, I do. Do you see that?<br>There [incomprehensible] are being distributed. | D points to the monitor.<br>C starts to type "while supplies are being distributed" |
| 11 | C | "while [incomprehensible] are being distributed" | |
| 12 | D | [incomprehensible] Oh, crap!<br>[incomprehensible] No, no, it does not work somehow. How embarrassing.<br>What's happening? You are clicking correctly, aren't you? | C is struggling with positioning the cursor at the word that needs to be corrected<br><br>They continue to watch the video to the end |
| 13 | D | "Joyful music, because the supply with food is ensured." | D moves the head in the rhythm of the music. |
| 14 | C | No, that's crap. | |
| 15 | D | So, what do you wanna write? „because [incomprehensible]"? | C types "Freudenmusik" [joyful music] |
| *Timestamp 00:25:12* | | *Experimenter announces that everybody should write their final comments and prepare to proceed to the following tasks.* | |

In conclusion, the results indicate that the digital video tools as "collaborative" tools supported students' understanding of historical sources in *different* ways during the visual design task.

## Discussion

Taken together, our quantitative and qualitative results indicate that advanced digital video technologies applied to classroom learning can be useful in supporting the socio-cognitive processes of student dyads performing complex visual design tasks in the history domain. The findings suggest, when explicit instructional guidance is limited, technological affordances and constraints can implicitly guide and support students' task-related and socio-cognitive activities. We found indications that the implicit instructional support of specific tools constitutes an important complement for learning with design tasks. However, students focused on applying the different technologies, but spent little time on planning and evaluation during task completion. This latter finding might explain why using instructional metaphors did not prove to be salient enough when combined with different digital video tools on students' learning: Metaphors are important during planning. When planning is limited, metaphors have limited impacts on design and learning.

The results of this field study are in line with the earlier findings from the lab study (Zahn, et al., in press), where the affordances of the advanced digital video tool impacted the quality of participants' design activities and extended to learning outcomes and learners' socio-cognitive processes focusing their discussions in very similar ways. Thus, we conclude that the effects of implicit guidance by technological affordances and constraints may constitute an important factor in computer-supported learning that needs to be considered with regard to educational goals and possible instructional strategies. Affordances and constraints of advanced technologies already in use can be seen as potentially supportive to implicit instructional aspects. Studying them means providing valuable information for effective integration in the classroom. However, explicit instructional aspects need to be considered, too, during learning with digital video. This is especially true when addressing issues such as optimizing learners' problem solving within the joint problem space of a complex visual design task, and successful integration of these tasks into instruction for a whole classroom. Further research is needed into these issues. In an initial step to look further into the role of explicit instructional support, we went back to a more controlled environment by inviting 201 students from 9 school classes to our research lab. In this study we manipulated two between-subjects factors: The first factor was the instructional support of either the design problem or the collaborative aspects of the task, which was derived from the theoretical model of design as problem solving. At the beginning of the collaboration phase, students were either prompted to consider the future audience of their design product and to make clear the main messages they wished to convey with their reworking of the source video (support of design), or they were asked to write down rules to ensure good communication within their dyad and to think about how to divide the task into sub-tasks and how to distribute these among them (support of collaboration). For the second factor, we manipulated the video tools, by contrasting two web-based applications: WebDIVER™ and Asterpix (http://www.asterpix.com). This manipulation was based on our considerations that we had found different technological systems to differ on a generic level in enabling users to acquire different cognitive skills (Zahn, et al., 2005): Some digital video technologies focus on integrating associated information, like the Asterpix system. Others focus more on observation and analysis, like WebDIVER™. We were especially interested in whether the specific socio-cognitive functions of the two prototypic tools would differentially influence student learning on the three levels described above (cognitive impact, action-related outcomes, socio-cognitive effect). Furthermore, we wanted to investigate whether our explicit instructions fostered students' planning behavior, and, indirectly, other important sub-processes of our model. Finally, we wanted to learn more about the interaction between the instructions and the tool-specific affordances, in order to provide more evidence for the need to frame digital video tools when integrating them into classroom instruction. Results from this study indicate significant effects for explicit instructional support, as well as the interaction with implicit guidance by the tools' affordances.

Taken together, such results provide initial evidence for the importance of instructional framing of visual design tasks. With our research we hope to deepen our understanding of how to foster students' video-related learning activities, dialogue, knowledge construction, or critical thinking when working with historical sources. We hope, too, to pave the way for constructivist approaches to using audiovisual media in the classroom.

## References


Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, *12*, 307-59.

Beichner, R. J. (1994). Multimedia Editing to Promote Science Learning. *Journal of Computers in Mathematics and Science Teaching*, *13*, 147-62.

Erickson, J., & Lehrer, R. (1998). The Evolution of Critical Standards As Students Design Hypermedia Documents. *Journal of the Learning Sciences*, *7*, 351-386.

Goldman, R. (2004). Video perspectivity meets wild and crazy teens: a design ethnography. *Cambridge Journal of Education*, *34*, 157 - 178.

Goldman, R., Pea, R., Barron, B., & Derry, S. J. (2007). *Video Research in the Learning Sciences*. Lawrence Erlbaum.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications*. (pp. 1-28). Mahwah, NJ: Lawrence Erbaum Associates.

Hobbs, R. (2006). Non-optimal uses of video in the classroom. *Learning, Media and Technology*, *31*, 35-50.

Jenkins, H., Clinton, K., Weigel, M., & Robison, A. J. (2006). Confronting the Challenges of Participatory Culture: Media Education For the 21st Century. Retrieved March 19, 2009, from http://www.newmedialiteracies.org/files/working/NMLWhitePaper.pdf.

Kafai, Y. B., & Resnick, M. (Eds.). (1996). *Constructionism in practice : designing, thinking, and learning in a digital world* (pp. XII, 339 S.). Mahwah, NJ: Lawrence Erlbaum Associates.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., et al. (2003). Problem-Based Learning Meets Case-Based Reasoning in the Middle-School Science Classroom: Putting Learning by Design-super(TM) Into Practice. *Journal of the Learning Sciences* , *12*, 495-547.

Krammer, R. (2006). Filme im Geschichtsunterricht - Analysieren-Interpretieren-Dekonstruieren. [Films in history classes - analyzing, interpreting, deconstructing.] *Historische Sozialkunde*, *3*, 26-33.

Lehrer, R., Erickson, J., & Connell, T. (1994). Learning by designing hypermedia documents. *Computers in the Schools*, *10*, 227-254.

Pea, R. (2006). Video-as-Data and Digital Video Manipulation Techniques for Transforming Learning Sciences Research, Education, and Other Cultural Practices. In *The International Handbook of Virtual Learning Environments* (pp. 1321–1393). Springer.

Pea, R. D., Mills, M. I., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert, E. (2004). The DIVER™ Project: Interactive Digital Video Repurposing. *IEEE Multimedia*, *11*, 54-61.

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*, 647-658.

Scardamalia, M., & Bereiter, C. (2006). Knowledge Building: Theory, Pedagogy, and Technology. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 97-118). New York: Cambridge University Press.

Smith, B. K., & Blankenship, E. (2001). Justifying imagery: Multimedia support for learning through explanation. *IBM Systems*, *39*, 749-767.

Spiro, R. J., Collins, B. P., & Ramchandran, A. (2007). Reflections on a Post-Gutenberg epistemology for video use in ill-structured domains: Fostering complex learning and cognitive flexibility. In R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 93-100). Mahwah: Lawrence Erlbaum Associates.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology*, *31*, 24-32.

Stahl, E., & Bromme, R. (2004). Learning by Writing Hypertext: A Research Based Design of University Courses in Writing Hypertext. In *Effective Learning and Teaching of Writing* (pp. 547-559). Retrieved July 10, 2009, from http://dx.doi.org/10.1007/978-1-4020-2739-0_37.

Wineburg, S. (2001). Historical thinking and other unnatural acts: Charting the future of teaching the past. Philadelphia, PA: Temple University Press.

Zahn, C., Krauskopf, K., Hesse, F. W., & Pea, R. (in press). Digital video tools in the classroom: How to support meaningful collaboration and reflective thinking of students. In M. S. Khine & I. M. Saleh (Eds.), New science of learning: Cognition, computers and collaboration in education. New York: Springer.

Zahn, C., Krauskopf, K., Hesse, F. W., & Pea, R. (2009). Participation in knowledge building "revisited": Reflective discussion and information design with advanced digital video technology. In C. O'Malley, D. Suthers, & A. Dimitracopoulou (Eds.), Computer Supported Collaborative Learning Practices: CSCL2009 Conference Proceedings (pp. 596-600). Brunswick, NJ: International Society of the Learning Sciences (ISLS).

Zahn, C., Pea, R., Hesse, F. W., Mills, M., Finke, M., & Rosen, J. (2005). Advanced video technologies to support collaborative learning in school education and beyond. In T. Koschmann, D. Suthers, & T. -. W. Chan (Eds.), *Computer supported collaborative learning 2005: The next 10 years* (pp. 737-742). Mahwah, NJ: Lawrence Erlbaum Associates.

Zahn, C., Pea, R., Hesse, F. W., & Rosen, J. (in press). Comparing simple and advanced video tools as supports for complex collaborative design processes. *Journal of the Learning Sciences.*

## Acknowledgments

This work was partly supported by grants from the Deutsche Forschungsgemeinschaft [German Research Foundation] (DFG ZA 524/1-1) and the National Science Foundation (NSF REC 0354453). Thanks to Joe Rosen and Benjamin Klages for helping during conducting the experiment and analyzing data. DIVER, WebDIVER, and Guided Noticing are trademarks of Stanford University for DIVER software and services with patents awarded and pending.

# THE EPISTEMOGRAPHY OF JOURNALISM 335: COMPLEXITY IN DEVELOPING JOURNALISTIC EXPERTISE

David Hatfield, David Williamson Shaffer, University of Wisconsin-Madison, 1025 West Johnson Street, Madison, WI 53706
Email: dhatfiel@wisc.edu, dws@education.wisc.edu

**Abstract:** As bloggers and mobile phone eye-witnesses increasingly supplement the 'news,' it is more important than ever to understand how professional journalists develop their expertise. In this paper, we examine an intermediate level reporting practicum course to explore the learning processes therein. Using a new method called Epistemic Network Analysis, we also explore emergent relationships within developing journalistic expertise. Understanding these relationships should be useful for journalism education as well as the design of research on learning environments.

## Introduction

Increasingly, bloggers and mobile phone eye-witnesses seem to provide the public with all the news that's fit to see. At the same time seemingly each year, reports (McDonnell et al., 2008) document the decline in newspaper readership, speculating that for the few people who still care about the news, Google and other online services are increasingly the destinations of choice.

We argue that despite commercial pressures and in light of the explosion of online information technologies, it is more important than ever to understand how professional journalists develop their expertise. It matters because that expertise is focused on the work Kovach and Rosenstiel (2001) describe as journalism's primary purpose: "to provide citizens with information," gathered and produced through "a discipline of verification," necessary "to be free and self-governing."

In this paper, we examine one of the ways journalists develop expertise through an ethnographic study of Journalism 335, an intermediate level reporting practicum course at a large Midwestern university. The main goal of the study was to explore the learning processes experienced by the students in the course. Using a new method adapted from social network analysis called epistemic network analysis (Shaffer et al., 2009), we also explore emergent relationships within developing journalistic expertise. Understanding these relationships should be useful for journalism education as well as the design of research on learning environments.

## Theory

In the professional practicum, novice professionals engage in simulations of professional work. Guided by reflective interactions with more experienced mentors, they "learn by doing," developing the "complex ensemble of analytic thinking, skillful practice, and wise judgment upon which each profession rests" (Sullivan, 2005, p.195). As Schon (Schon, 1983, 1987) argues, for most professions this "complex ensemble" is developed in capstone courses and/or professional practicum experiences.

Shaffer (2004, 2005a), extending Schon, calls this ensemble an *epistemic frame*, emphasizing that what makes professional expertise uniquely powerful are the relationships between the elements within the ensemble. In other words, the particular set of knowledge and skills that is the focus of most "hands-on" approaches is interconnected with a particular set of values and identities, forming a coherent and integrated perspective that professionals use when making professional judgments in the world. Thus, a professional journalist has mastered a body of writing and reporting knowledge and skills, sees herself and is seen by others as a journalist, and frames her efforts according to a set of journalistic values and norms.

Shaffer (2005b) argues that the practicum is designed to help develop this professional epistemic frame through its participant structures, or the "recurrent patterns of involvement that structure a particular kind of situation within a given practice." In his study of Journalism 828, an advanced reporting capstone course, Shaffer (2005b) explores three such participant structures: interactive copy editing of stories with peers and the professor, news meetings in which participants shared progress and received feedback on their reporting, and war stories from the professor or guest journalists providing lessons learned from past experience. Each participant structure "links identity and practice with particular knowledge and values of a domain" (Shaffer, 2005b) for one of three professional abilities: writing to formula, writing for story, and writing as a watchdog.

The ability to write to formula reflects the highly prescribed nature of journalistic writing (Shaffer, 2005b). From capitalization and punctuation rules in the Associated Press (AP) style guide to innumerable how-to guides for story organization and structure, being able to appropriately use formula is critically important (Murray, 2000; Edgerton, 1997). But this skill is also bound with knowing a particular set of principles concerning appropriate usage and application of this skill. It also involves adopting a particular kind of voice, or identity, within the writing as a way of asserting certain values of neutrality, objectivity, if not truth (Kovach & Rosenstiel, 2001; Stewart, 1998).

Writing for story, on the other hand, focuses on the practices of reporting and developing the particulars of stories in order to realize the broader significance of specific characters and details (Shaffer, 2005b; Franklin, 1986). It employs knowledge about various reporting techniques and tools, skill in their use to capture key details, as well as knowledge and skill with narrative techniques to most effectively use those details in stories. It emphasizes the reporter's identity as a "professional pest," tenaciously seeking out information, and the value of letting the voices of real people carry the story.

Finally, writing as a watchdog organizes the various elements involved in the journalist's role as "watchdog for the public trust" (Shaffer, 2005b). This involves developing skills to "monitor power and offer voice to the voiceless" (Kovach and Rosenstiel, 2001) through knowledge of the normal activities of institutions and individuals, both powerful and humble, and it is guided by the profession's ethical value of public accountability.

Each of these abilities thus constitutes its own "complex ensemble" within the broader professional frame of journalistic expertise. They constitute, in other words, epistemic *sub-frames* in which particular kinds of knowledge, skill, identity, values and epistemology are bound together, and which should also feature linkages to other sub-frames as part of an overall integrated epistemic frame.

Shaffer (2005b) argues that each ability is developed separately and then integrated with the others through the epistemic discourse of the professor, as for example in discussions of how the different practices constitute "smart reporting." In this study, we explore an alternative possibility, that these abilities may each exhibit different patterns of internal development and integration with the others, and we describe a technique for quantifying these patterns called the Integration-Cohesion index below.

Here, we extend Shaffer's study by examining a similar journalism education context to investigate whether these three sub-frames manifest within a transitional practicum course. We then use epistemic network analysis to examine how these sub-frames become integrated into a unified professional perspective. I argue that epistemic network analysis can provide a quantitative measure of the complex linkages within and between ensembles of knowing, being, acting and valuing that comprise a professional perspective. Finally, I discuss how the results of this study may contribute to the design of learning environments.

## Methods

### Setting

As part of the University's School of Journalism and Mass Communication "Reporting" track, J335 occupies a key transitional role between highly structured introductory courses and more open advanced writing and reporting courses, thus providing an opportunity to study an early practicum environment.

J335 was a four-credit course that met twice a week for 75 minutes throughout a 15-week semester. Students also met outside of class in small groups while working on two different group reporting projects, the first occurring midway through the semester and the second occurring at the end of the semester. The specific sequence of assignments is described in the results section.

### Participants

The class was comprised of 13 students. There were six seniors and seven juniors, all journalism majors. No other demographic information was collected about the students. The professor, John, is a prize-winning journalist. The class also had five guest speakers, each of whom was a reporter, editor, or publisher at local or regional newspapers or radio organizations.

### Data collection

Data was collected about the course in several ways. Twenty six of the 30 class meetings were observed. During the semester, students worked on two team reporting projects, and one group was observed for the second project. This team included four seniors and one junior. Copies of all e-mail messages sent by the professor to the class and also of messages exchanged among members of the project group were collected and analyzed. Copies of all story assignments including both the story text and the professor's feedback were also collected and analyzed.

Classroom data was collected in digital audio recordings and supplemented with field notes. Recordings were transcribed to provide a detailed record of interactions, and field notes were used to capture meaningful non-verbal aspects of the context and supplement the transcripts. Out-of-class group meetings were documented via field notes.

## Data analysis

The data was segmented into interactive units, which were defined as strips of activity with a consistent interactional structure and topical focus. The boundary between strips was marked by either a change in topic or a change in interactional structure. In other words, when either the focal point of discourse or the social arrangement changed, one unit ended and another began. For in-line story feedback, comments targeting the same passages in a story were treated as occurring within a single unit while comments targeting different parts of a story were treated as occurring in separate interactive units. Specific examples of classroom interactive units are provided below in the results section, but in the class the transition was clear and often explicitly marked by professor comments like "Alright, we're going to change gears now and focus on [something new]."

Strips were coded for particular practices, knowledge, identities, and values of each of the three epistemic sub-frames of entry-level journalism identified previously: 1) writing to formula; 2) writing for story; and 3) writing as a watchdog. Following Shaffer (2005b), segments were coded for writing to formula when participants "discussed or referred explicitly to the methods of journalistic writing, including specific rules of journalistic style and forms or formulae of traditional journalism." Segments were coded for writing for story when participants "discussed or referred explicitly to journalism as a process of telling stories: stories about particular people encountering problems or conflict, the thoughts and actions of those people, and the specific events that happened and the unique details surrounding them." Segments were coded for writing as a watchdog when participants "discussed or referred explicitly to informing the public about important information and events, drawing attention to inequities, monitoring people and institutions in positions of power, or offering a voice to those without power." Table 1 provides a brief description and example for each category used in this analysis. If a strip represented more than one category, it was coded for all applicable categories.

Table 1: Coding scheme for observations and stories using the elements of an epistemic frame

| Code | Description | Example |
|---|---|---|
| Writing to Formula | | |
| Practices | Using journalistic writing formula | "A terrific job of drawing readers into the lead by setting the scene and mixing a little action with quotes." |
| Identity | Invisible authorship | "No need to state your conclusions in the ending. Just let your sources and docs carry the story." |
| Values | Writing with a neutral voice | "Interviews where you speak the most will probably be the most fun and the most worthless." |
| Knowledge | Knowledge of journalistic style | "Please review AP style; his title wouldn't be capitalized unless it immediately precedes his name." |
| Writing for Story | | |
| Practices | Skills of smart reporting | "Couldn't we get the police report and search warrant return to give readers this powerful piece of info?" |
| Identity | Identity as a professional pest | "Part of your job is to assess credibility. It's not just to get that good quote, but it's to aggressively assess whether your source really knows the issue." |
| Values | Importance of providing an individual face | "Strive to use character development, description, narrative storytelling and use of key details and facts to grab readers' attention. For example, the story could focus on John Doe with a little additional reporting." |
| Knowledge | Knowledge of reporting tools | "There's more information more readily available to journalists today than ever in history. … The challenge is knowing which kinds of information to acquire and what to do with it." |
| Writing as a Watchdog | | |
| Practices | Recognizing important patterns | "You found ways to bring readers inside the world of students worrying about being admitted …. And officials shared insights into how things are supposed to work." |
| Identity | Acting as a watchdog | "Would be even better with interviews with some of the candidates and election officials. Still, this is an important issue and I applaud you for tackling it." |
| Values | Importance of public accountability | "Given the recent revelations about this charity's financial woes, would be important to include a graph or two about that." |
| Knowledge | Knowledge of journalistic ethics | "a good attempt to use data. But wouldn't it be far more fair to compare the campus ethnic mix to that of the state and city rather than the U.S. as a whole?" |

The result was a database of interactive units showing the presence of epistemic frame elements organized by epistemic sub-frames for entry-level journalism as well as the times during which they occurred. The relationships among these different components were then analyzed using epistemic network analysis.

### Epistemic network analysis

Epistemic Network Analysis (ENA) adapts the framework of Social Network Analysis (SNA) for use with cognitive, rather than social, elements (Shaffer et al., 2009). As Shaffer argues, while SNA was developed to provide insight into the relationships between individuals and groups, it also provides a robust set of analytic tools for representing and studying networks of relationships in different domains. In ENA, these tools can be useful for understanding the kinds of linkages between and across elements taking place over time.

For this study, each epistemic frame element was considered to be a node in the epistemic network representing the interaction captured in each interaction unit or strip. Links were defined as the co-occurrence of two or more epistemic frame elements within any strip. Links between the elements of a single sub-frame were defined as intra-frame links, while links from elements in one sub-frame to those in another were defined as inter-frame links.

In order to capture the strength and changes of these links over time, we developed an Integration-Cohesion (IC) index. The IC index measures the degree to which a given sub-frame or sub-frame element emphasizes cohesion, by linking more often to elements in the same sub-frame, or integration, by linking more often to elements of other sub-frames (see Table 2).

Table 2: Intra-frame and Inter-frame linkage formulae

| | |
|---|---|
| $Intra-frame = \sum \sqrt{(L_{in})^2 / (M_{in}-1)}$ | $Inter-frame_{sub1} = \sum \sqrt{(L_{sub1})^2 / (M_{sub1})}$ |
| Equation 1: intra-frame linkage | Equation 2: inter-frame linkage |

As provided in Equation 1, the strength of intra-frame linkage for any frame element was calculated by first summing the squared values of the links between that element and the other frame elements in that element's sub-frame ($L_{in}$), then dividing by the number of elements in the sub-frame ($M_{in}$) minus one, and finally taking the square root of the resulting number. As provided in Equation 2, the strength of inter-frame linkage for any frame element was similarly calculated by first summing the squared values of the links between a given frame element and the frame elements in one of the other sub-frames ($L_{sub1}$), then dividing by the number of elements in that other sub-frame ($M_{sub1}$), and finally taking the square root of the resulting number. The inter-frame calculation was repeated for each of the two sub-frames in which each frame element was not a member. These values were then summed at each successive interaction unit, producing one intra-frame and two inter-frame linkage values which reflected the accumulated link behavior to that point in time. Finally, the ICI was calculated as the difference between the inter-frame linkage value and the intra-frame linkage value at each successive interaction strip, with positive values indicating integration across two different sub-frames and negative values indicating cohesion within a single sub-frame.

## Results

Journalism 335 (J335) was an intermediate reporting class designed, according to the syllabus, "to develop reporting and writing skills." As a pre-requisite for several advanced writing courses, this course provides "practical training in a wide range of reporting techniques, including interviewing, use of public records and research methods" to help students "learn how to become better watchdogs over powerful individuals and institutions and how to provide a voice for the voiceless."

Our observations of J335 are described below. We describe the epistemic sub-frames, *writing to formula*, *writing for story*, and *writing as a watchdog*, that emerged during the course. Then, for each sub-frame we describe the Integration-Cohesion index relationships that manifested between the sub-frames.

### Epistemic sub-frames

Shaffer (2005b) argues that journalism reporting courses are organized around at least three "critical elements in the skill set of an entry-level journalist: the ability to write to formula, the ability to write for story, and the ability to write as a watchdog" (p12). In turn, each of these three abilities comprises an interlinked and integrated set of particular journalistic knowledge, skill, value and identity relative to that ability. These three epistemic sub-frames played an important role in organizing the experiences of students in J335 as well.

## Writing to formula

As the syllabus suggested, part of the emphasis in J335 is to help students develop journalistic writing skills. In one interactive unit during class, for example, John and the students discussed different kinds of leads and when it is appropriate to use them in stories. John asked, "Blind leads, where are they helpful?" Carol replied, when the news is "important but people don't know the organization." "Right. The issue is interesting, but the group isn't well known," John said,

> "Something should be going off in your mind right at that point saying, maybe this is the right structure for this kind of story. Now generally if you do go with the blind approach, that's going to also suggest to you very quickly usually within a couple sentences, you'd better circle back in and start filling this in with some of that detail you omitted from that first graph cause you don't want to leave your reader hanging too long."

In this brief excerpt, John showed the students a particular journalistic writing *practice*: when "the issue is interesting but the group isn't well known," a blind lead may be "the right structure for this kind of story." He cautioned about a potentially conflicting journalistic writing *value*: "you don't want to leave your reader hanging too long." And he used specific journalistic writing *knowledge* to suggest ways of avoiding this problem, by "circling back in and filling this in with some of that detail you omitted from that first graph." In other words, he articulated an interconnected sub-frame of values, practices and knowledge journalists use to *Write to Formula*.

Using epistemic network analysis (ENA) to quantitatively look across all of the strips of activity for the semester as shown in Figure 2, the *Writing to Formula* sub-frame demonstrated relatively strong internal cohesion, particularly for its epistemic frame knowledge and skill elements. In other words, interactions throughout the course worked to develop rich intra-connections within this sub-frame, focusing in particular on binding understanding of journalistic writing formula with skill at its use.

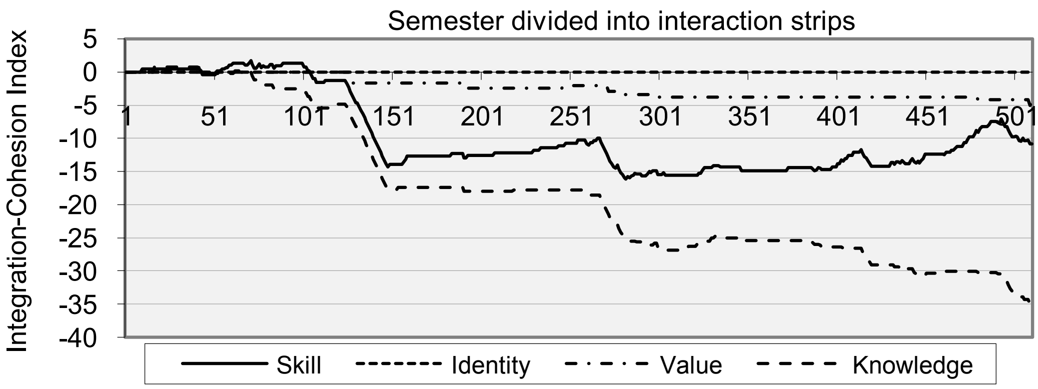



Figure 1. Writing to Formula Integration-Cohesion Index

## Writing for story

As a reporting class, J335 was equally concerned with helping students learn to develop reporting skills. In an example from another class session later in the semester, John was reviewing Susan's education story in front of the class. "One thing I liked about this story," he commented

> "was instead of just saying it's a poor school district, there were some good supporting numbers. [Median household income for the town, median for North Carolina.] … There's also a description - the interior is shabby … Then we get down here, this different view that visitors would see at Randall Elementary. … So you get some idea of the disparity in what they look like as well as then disparities in the teachers."

Susan responded, "that was one of the interesting things about going to North Carolina. They were saying that they were taking teachers that weren't even certified. … I assumed that you needed to be certified." John then wrapped up the review by commenting,

> "one thing that I think works here that is a good story telling device that you could think about for your own future work is, if you have a visual element that you think would help people understand and you have some important issue like the qualifications of teachers which is kind of an abstract concept, think about starting with the visual because that's more comprehensible, more tangible for your readers to describe things and point out disparities, visual disparities. Then when you've softened people up and you've gotten them interested in it, then talk about the disparities with teachers. … And in your own work, just keep looking for disparities like this. You can never go wrong by looking at situations involving the have's and the have-not's, whether you're talking about education, housing, whether you're talking about safe neighborhoods or unsafe neighborhoods. And these are issues that are ripe for journalists too ... Because they are open to us, few institutions point these things out, so that's part of our job is to explore these situations."

In this example, John showed the students a particular reporting *practice*: gathering and using "good supporting numbers," the median household income for the two communities in the story. This practice was guided by a particular reporting *value*: "You can never go wrong by looking at situations involving the have's and the have-not's," and connected to a particular reporting *identity*: "that's part of our job is to explore these situations." In other words, he highlighted the connections between a particular sub-frame of values, practices and identity journalists use to *Write for Story*.

Similar to the *Writing to Formula* case, using ENA to look across all of the semester's strips of activity, as shown in Figure 3, reveals developing internal cohesion for the epistemic frame elements of the *Writing for Story* sub-frame. In addition, the cohesion shown for this sub-frame is stronger than that for *Writing to Formula*, as shown by the relatively larger negative movement in the I-C index, and all of the elements are involved. Finally, the identity element of *Writing for Story* was also substantially more cohesion oriented in contrast to identity in *Writing to Formula*. These data suggest again and perhaps even more strongly that interactions throughout the course worked to develop rich intra-connections within this sub-frame, mobilizing the reporter's identity as an interviewer and storyteller along with developing skills and knowledge of techniques.

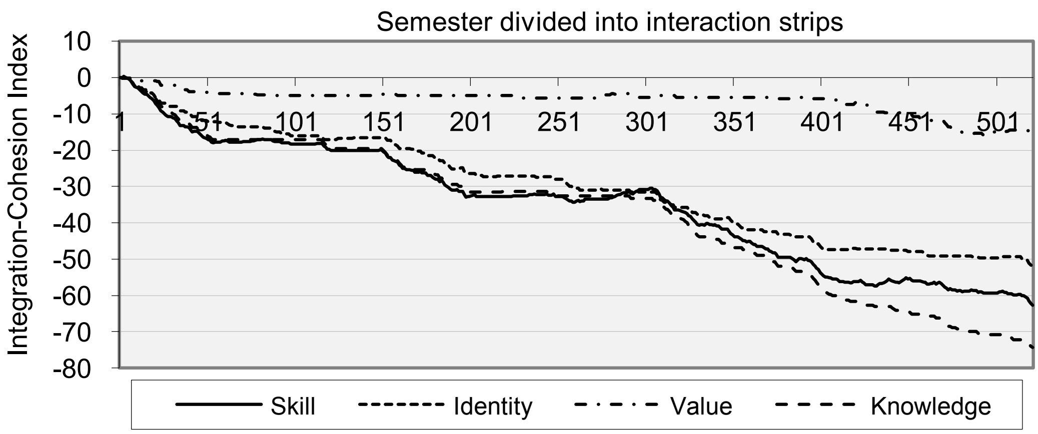



Figure 2: Writing for Story Integration-Cohesion Index

## Writing as a watchdog

In J335, students were developing reporting and writing skills in order to "learn how to become better watchdogs over powerful individuals and institutions and how to provide a voice for the voiceless." After taking the class to meet with the County Clerk, the Police Department spokesperson, and the Mayor, John and the class discussed their interactions with these powerful institutions. One of the students, Sean, said, "as far as what [the Police spokesperson] was saying about be accurate and fair, seemed he stressed be nice to them. Isn't there a point where you don't want to be so accommodating to them? … His whole message to us was 'be nice to us and do what we tell you to do.'" John replied,

> "Right. And through your whole career you're going to get that attitude, especially from people in law enforcement and paramilitary organizations – they're used to dealing with

> hierarchy. Journalism breaks down hierarchy. Journalists seek to understand hierarchies, and then seek to penetrate those hierarchies. People in charge of those hierarchies often resent those efforts. You're exactly right-there's a tension there. I think the challenge for us as journalists is how to achieve the balance. It's possible to be both polite and aggressive. It's possible to be nice but yet firm. To articulate our point of view in a respectful manner, but keeping in mind that we have a different point of view and a distinctly different role than a public agency does. Our responsibility should be to whom? Who are we serving? The public, right."

In this case, John showed the students how a particular watchdog *identity* was connected to a particular set of watchdog *practices*: "Journalists seek to understand hierarchies, and then seek to penetrate those hierarchies." This linked identity and set of practices required a particular watchdog *understanding* guided by a particular watchdog *value*: how to achieve a balanced articulation of the journalist's point of view "in a respectful manner" while keeping in mind "our responsibility" to the public. In other words, he focused on the particular ways of understanding and valuing that guide journalistic practice and perspective when journalists *write as a watchdog*.

The ENA picture that emerges for the *Writing as a Watchdog* sub-frame, shown in Figure 4, is quite different from the other two sub-frames. In this case, all of the elements are revealed to be linking either neutrally (neither more internally nor externally linked) or strongly externally as in the case of the skill element, integrating this sub-frame with the others. These data suggest that interactions throughout the practicum used *Writing as a Watchdog* elements, and particularly the practices of pattern recognition central to this sub-frame, to tie together the developing sub-frames of formulaic writing and story reporting.

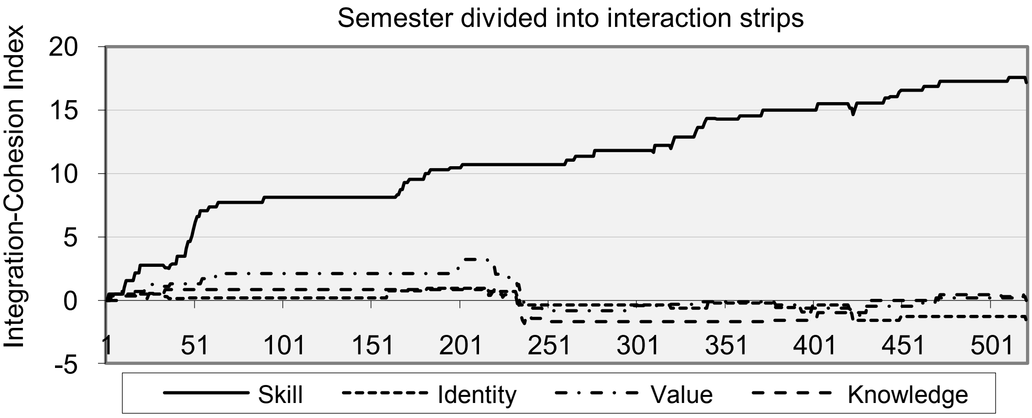



Figure 3: Writing as a Watchdog Integration-Cohesion Index

As described earlier, Shaffer's epistemography of the advanced reporting capstone course Journalism 828 (Shaffer, 2005b) provides a compelling account of the development of these abilities through the learning practices of the Journalism 828 practicum. While not surprising, the results of this study do suggest that these epistemic sub-frames organized student experience throughout the course in ways quite similar to those of the advanced practicum course. In addition, the ENA data provides a new perspective on the ways in which pedagogy was instantiated on the ground in course interactions

## Discussion

These results suggest that over the course of Journalism 335, distinct elements of journalistic expertise were drawn out and linked together, helping to develop and integrate the epistemic sub-frames of *writing to formula*, *writing for story*, and *writing as a watchdog*. This confirms similar observations in Shaffer (2005b), suggesting that, despite being a transitional course, J335 provided an important practicum experience for its journalists in training.

In addition, the results of this study indicate that different sub-frames for specific areas of expertise can exhibit different developmental patterns, sometimes emphasizing the development of stronger connections with other elements of the same sub-frame, and sometimes emphasizing stronger connections with elements from other sub-frames. This in turn suggests an area for future study might be exploring the ways in which developers of learning environments might best align different kinds of participant structures to take advantage of the different developmental patterns touched on here.

These results also suggest that the epistemic network analysis method can be productive way of studying interactions within learning environments. Building on Shaffer (2009), this study adds the *Integration-Cohesion* index to epistemic network analysis, to uncover particular linkages patterns occurring across sometimes difficult to see interactions. Epistemic network analysis suggests a powerful set of techniques for studying the traces of cognitive development through social interaction, and it points toward a promising new way of observing the translation of pedagogy into practice in various kinds of learning environments.

The study presented is of course limited. As observations of a single seminar, these results obviously can't generalize to all instances of journalism practica or capstone courses. Further, as an intermediate-level course, there was significantly less student-initiated interaction than might have been observed in more advanced courses. At the same time, these findings do suggest a more complex relationship within and between these epistemic sub-frames can emerge during the overall development of journalistic expertise, the details of which deserve future study. They also suggest that the method of epistemic network analysis may be useful for future investigations of learning environment interactions for a deeper understanding of cognitive development.

## References


Edgerton, L. (1997). *The Editing Book*. Dubuque, IA: Kendall/Hunt.

Franklin, J. (1986). *Writing for story: craft secrets of dramatic nonfiction by a two-time Pulitzer Prize winner* (1st ed.). New York: Atheneum.

Hatfield, D. & Shaffer, D.W. (2006). Press play: designing an epistemic game engine for journalism. Paper presented at *The 2006 International Conference of the Learning Sciences (ICLS)*. Bloomington, IN.

Kovach, B., & Rosenstiel, T. (2001). *The elements of journalism: what newspeople should know and the public should expect* (1st ed.). New York: Crown Publishers.

McDonnell, J., Mitchell, A., & Rosenstiel, T. (2008). The State of the news media 2008. Project for excellence in journalism. Retrieved on May 28, 2008, from: http://www.stateofthenewsmedia.org/2008/index.php

Murray, D. M. (2000). *Writing to deadline: the journalist at work*. Portsmouth, NH: Heinemann.

Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Schon, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.

Shaffer, D. W. (2004). Epistemic frames and islands of expertise: Learning from infusion experiences. In Y. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon & F. Herrera (Eds.), *Proceedings of the Sixth International Conference of the Learning Sciences* (pp. 473-480). Mahwah, NJ: Erlbaum.

Shaffer, D.W. (2005a). Epistemic Games. *Innovate 1*(6). Reprinted in *Computer Education* (in press).

Shaffer, D. W. (2005b). Epistemography and the Participant Structures of a Professional Practicum: A story behind the story of Journalism 828. *(WCER Working Paper No. 2005-8)*: University of Wisconsin-Madison, Wisconsin Center for Education Research.

Shaffer, D. W., Hatfield, D. L., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E. A. S., et al. (2009). Epistemic Network Analysis: A prototype for 21st Century assessment of learning. *The International Journal of Learning and Media, 1*(2), 33-53.

Stewart, J. B. (1998). *Follow the story*. New York: Simon & Schuster.

Sullivan, W. (2005). *Work and Integrity: The Crisis and Promise of Professionalism in America*. San Francisco, CA: Jossey-Bass.

# Which science disciplines are pertinent? –Impact of epistemological beliefs on students' choices

Porsch, Torsten & Bromme, Rainer, University of Muenster, Fliednerstraße 21, 48149 Muenster,
Email: t.porsch@uni-muenster.de, bromme@uni-muenster.de

**Abstract:** The growth of scientific and technological knowledge in modern societies has lead to an increase of specialization of knowledge and expertise. Most socio-scientific issues are far too complex to be understood deeply by laypersons. From various disciplines we have to choose pertinent ones if we want to rely on expert advice. Epistemological beliefs might be helpful to cope with this challenge. Furthermore it is necessary to have realistic awareness of one's own fragmentary understanding and to avoid the "illusion of explanatory depth" (Rosenblit & Keil, 2002). In order to research on adults' capability to choose between disciplines who might be relevant for a science topic, N = 520 secondary school students were asked to choose, which of 22 scientific disciplines (e.g. math, geology, biology) should contribute to a book about tide and float. They were also asked to assess their own knowledge about the theme. Influence of epistemological beliefs has been tested by an epistemological sensitization in an experimental design. The epistemological sensitization significantly influences students' self-assessment of knowledge and discipline rating. Students with sophisticated epistemological beliefs were more critical about their own knowledge about tide and flow, chose significantly pertinent and -by tendency- potential pertinent disciplines more and declined non-pertinent disciplines more.

## Introduction

### Division of cognitive labor

The enormous growth and importance of scientific and technological knowledge in modern societies entail various consequences. Knowledge is unevenly distributed and we are laypersons in most knowledge domains. Dealing with various problems demand different and diversified knowledge backgrounds. Most socio-scientific issues (chemical additions to food, genetic engineering; Zeidler, Sadler, Applebaum & Callahan, 2009) are far too complex to be understood deeply by laypersons, but it is nevertheless necessary for laypersons to reason and to decide about these issues with regard to personal health or with regard to civic participation. Laypersons must rely on experts based on their own fragmentary understanding of such issues (Bromme, Kienhues & Porsch, 2009). Even children learn from the very beginning that they have to rely on others in order to get information, to answer a question, or to solve a problem (Bergstrom, Moehlmann, & Boyer, 2006). Children know that they can ask parents and peers if they need to know something. They differentiate who to trust (Harris, 2007), they are aware that people have different areas of expertise, and know how knowledge is clustered in the minds of others (Keil et al., 2008). For instance, in a study by Lutz and Keil (2002), 56 children of different age (3-, 4-, and 5-year-olds) had to judge which of two experts (a car mechanic or a doctor) would know more about a specific topic. Results show that the 4-and 5-year olds were able to attribute knowledge correctly based on scientific principles, whereas the 3-year olds performed better than chance merely on stereotypical role items. Even children have an early notion of the division of cognitive labor, they have an assumption which experts might know something about what. In the following we will call this knowledge about the "disciplines` pertinence" to certain issues.

An understanding of the pertinence of different knowledge fields to different issues must be based on an understanding of the structure of knowledge. Scientific knowledge has been developed in and is taught in academic disciplines, for example in school or at universities (Stevens, Wineburg, Herrenkohl, & Bell, 2005). We all grow up with and learn knowledge being structured into different disciplines. But not all issues and problems could be clearly and unambiguously assigned to one and only one discipline. Some issues are multi disciplinary and do not belong to only one discipline. Especially new, emerging problems and complex socio-scientific issues fell into the domain of several disciplines. Furthermore, the academic canon of disciplines is subject to continuous change. While it would be desirable if layperson would know who (which expert?) would be pertinent to which issue, it can be difficult asses which disciplines are more or less relevant for a theme. It is an open empirical question if secondary school students are able to identify disciplines which are pertinent to a complex scientific issue.

### Epistemological beliefs

Research on epistemological beliefs, i.e. beliefs about the nature of knowledge and knowing, has expanded considerably in recent years (see, for overviews, Buehl & Alexander, 2001; Hofer & Pintrich, 1997, 2002). Epistemological beliefs include beliefs about the structure and variability of knowledge (Stahl &

Bromme, 2007). We can assume that someone conceiving knowledge as static, (if it has been found once it will not be changed anymore) and as a collection of clear, but separated facts that could be “found”, would have less elaborated ideas about the pertinence of disciplines to issues. Naïve or better to say “straightforward” epistemological beliefs might come along with a less elaborated awareness about the possibility to have different kinds of knowledge about the same issue (object of reference). We assume that such a person might have less well elaborated ideas about the pertinence of disciplines to topics, because she or he might tend to identify the knowledge with its objects of reference. In contrast, a person holding more sophisticated beliefs about knowledge (i.e. conceiving knowledge as dynamic and interrelated and more constructed than “found” ) might have more elaborated ideas about the pertinence of disciplines to issues. Furthermore he or she might include more disciplines when asked which discipline might be pertinent to a certain topic. We expect therefore the following relationship between epistemological beliefs and judgments about the pertinence of disciplines to complex issues: Sophisticated epistemological beliefs go hand in hand with a better (more realistic) assessment of disciplines pertinence. Dealing with a complex scientific issue, sophisticated beliefs should foster the inclusion of various disciplines.

### Knowledge self-concept

In informal settings it is necessary to have realistic metacognitive awareness of one’s own fragmentary understanding and to avoid the “illusion of explanatory depth” (Rosenblit & Keil, 2002). It is an open question if the self-assessed knowledge about the topic impacts on layperson’s assumptions about the pertinence of disciplines to complex scientific issues. Such an impact could be all the more probable as students self concept about their own knowledge differs between different school subjects. Even more, research on students’ self concepts has shown that students tend to overestimate the differences between their knowledge and abilities within different school subjects (Marsh & Hau, 2004). These findings of self-concept research argue for a cognitive and emotional impact of the disciplinary structure of school subjects on academic self concept, and as we conclude, they point to the possibility that the assessment of one’s own knowledge is relevant for judgments about the pertinence of disciplines to topics.

There is broad evidence for gender differences in self-concept (e.g. Marsh & Craven, 1997; Jacobs et al., 2002). Dealing with knowledge in school context boys tends to have higher self-concepts in subjects like math, whereas girls show a higher self-concept in language arts. Dealing with a complex scientific issue with a main focus on natural science, male students are supposed to have higher knowledge self-concepts. It is an open empirical question if the self-assessed knowledge about a theme moderates the rating of discipline’s pertinence and if this is also influenced by gender.

### Research questions

In this study we address the following research questions: First, do students know which disciplines are pertinent? Based on the division of cognitive labor we expect that students are able to identify which disciplines are pertinent. Second, is the rating of discipline`s pertinence impacted by student’s epistemological beliefs? We expect that sophisticated epistemological beliefs foster the ability to assess discipline`s pertinence. Third, is the rating of discipline`s pertinence moderated by student’s knowledge self-concept? Furthermore, the influence of gender on knowledge self-concept and thereby also on pertinence ratings will also be considered.

## Method

### Procedure

Data collection was done on personal computers. During the session all students read an introduction text about tide and float. Students were randomly assigned to two sub-samples that received two versions of this introduction text about tide and float. Before and after reading the introduction text, an epistemological beliefs questionnaire has been conducted. Subsequently, students assessed their own knowledge about “tide and float”. In the main part of the study, students had to evaluate 22 disciplines regarding to their pertinence to the theme “tide and float”. The whole session took around 20 minutes for each student. Due to the computer supported session we have not to deal with missing data on the depending variables. When a question has not been answered, students have been pleased to do so before proceeding.

### Participants

Secondary school students were randomly recruited during an open day at the university ($N$ = 520). Mean age was around eighteen ($M$ = 17.95; $SD$ = 1.01). The majority of the participants was female (132 male, 382 female, 6 unstated).

### Materials

<u>Epistemological Beliefs Questionnaire</u>

The success of the epistemological sensitization was determined by administering an instrument based on a semantic differential. The instrument was administered before and after reading the introduction text. The CAEB (Connotative Aspects of Epistemological Beliefs; Stahl & Bromme, 2007) consists of two scales of connotative adjective pairs: CAEB-texture measures beliefs about the structure and accuracy of knowledge on 10 items (sample item: "structured – unstructured") and exhibited satisfactory reliability pre-instructionally (Cronbach's $\alpha = .78$) as well as post-instructionally (Cronbach's $\alpha = .85$). CAEB-variability measures beliefs about the stability and dynamics of knowledge on 7 items (sample item: "dynamic – static") and exhibited medium reliability preinstructionally (Cronbach's $\alpha = .64$) and satisfactory reliability post-instructionally (Cronbach's $\alpha = .83$). The dimensions "texture" and "variability" are similar to the factors simplicity, structure and certainty from the Hofer model (Hofer & Pintrich, 1997, 2002) and the "source" dimension as described e.g. by Schommer (1990), but are named and conceptualized partly different. Texture encompasses beliefs about the structure and accuracy of knowledge. This factor includes beliefs from the dimensions "simplicity" and "source", and from "certainty". Variability encompasses beliefs about the stability and dynamics of knowledge. This dimension ranged from beliefs that knowledge is dynamic and flexible to beliefs that it is stabile and inflexible. This dimension mainly encompasses the dimension "certainty" but also the dimension "source". Several studies have shown that the CAEB is a quite reliable and valid instrument (e.g. Stahl, Pieschl & Bromme, 2006; Pieschl, Bromme, Porsch & Stahl, 2008; Kienhues, Bromme & Stahl, 2008). For the measurement of epistemological beliefs in this study, it has been used in a topic related way, i.e. it had to be sensitive to the knowledge about the theme "tide and float" and students had to refer their answers just to their beliefs about knowledge from this theme. This takes into account, that such beliefs can be topic-specific (e.g. Trautwein & Lüdtke, 2007).

## Introduction text

Ideas about pertinence of disciplines can be related to more and less sophisticated epistemological beliefs (see above). To consider these claims experimentally it is necessary to manipulate epistemological beliefs systematically: Two versions of an introduction text to "tide and float" were randomly administered to the students as an instructional intervention that we termed epistemological sensitization. In a recent study it was possible to successfully sensitize subjects reading science texts for an advanced epistemological stance by a comparable method (Kienhues, Bromme & Stahl, 2008). The two sub-samples did not differ with regard to age, gender, level of education, and school performance (measured by grades). One sub-sample ($N = 258$) read an introduction which was "straightforward" because it was enriched with statements that indicate a more structured and static view on the knowledge about tide and float (e.g., "models about tide and float fit well"; "water level forecasts are accurate even in the long run"). The other sub-sample ($N = 262$) received an introduction text that was enriched with comments on the epistemological nature of selected facts (e.g. detailing scientific controversies) and thus should elicit a more "sophisticated" view on the knowledge. Results show that the epistemological sensitization heavily impacted the ratings on the (Post-) CAEB-Questionnaire (repeated measure ANOVA for Texture: $F(1, 518) = 148.56$; $p < 0.00$, $\eta^2 = 0.22$ and Variability: $F(1, 518) = 301.94$; $p < 0.00$, $\eta^2 = 0.37$). The epistemological sensitization impacted the ratings into the assumed directions. For example, students in the straightforward instruction group rated their knowledge about tide and float as more structured and static in the Post-Questionnaire. In the sophisticated group students rated their knowledge as more unstructured and dynamic. Instruction groups are included as independent variables in subsequent analyses. For our research questions it is not important if this change is a fundamental and lasting modification of epistemological beliefs or a temporal effect on context-dependent epistemological resources (Hammer & Elby, 2003). We hereby acknowledge that our epistemological sensitization might only have changed epistemological beliefs during the session (in situ).

## Knowledge self-assessment

Students rated their own knowledge about "tide and float" on 5 Items (e.g.: "I already know very much about tide and float" or "There is not much more, that I can learn about tide and float" (reversed)). Students answered each item on a 5-point-Likert-scale (from "I totally disagree" to "I totally agree"). The scale exhibited satisfactory reliability (Cronbach's $\alpha = .84$). Students rated their knowledge above mean ($M = 3.08$; $SD = 0.88$). The individual results of the knowledge self-assessment scale are admitted in further analysis. Due to the above mentioned results, gender will be included as an independent variable in subsequent analyses.

## Discipline evaluation

To measure how students assess the pertinence to difference disciplines, the following scenario was established. We told the students that all the knowledge about tide and float is going to be gathered in a book. They had to decide which disciplines should contribute to this book. Each discipline had to be rated on a 5-point-Likert-scale (from "should absolutely not contribute" (1) to "should absolutely contribute"(5)). The result of the discipline evaluation is admitted in further analysis as a depended variable. Twenty two different disciplines were listed (randomized for every student): forestry, meteorology, geo science, life sciences, landscape ecology, information technology, electro technology, business informatics, e-business, information management, physics, chemistry, biology, math, medicine, orthopedagogy, museology, astronomy, oceanography, engineering, geology, and hydrology. The disciplines have been collected from the domains "environment", "general knowledge", "computer" and "health". There are some clearly pertinent disciplines (e.g. oceanography) but there are also cases where the relevance for the topic is less obvious.

# Results

## Do students know which disciplines are pertinent?

This question was investigated by using an external criterion. Disciplines` pertinence is not as obvious at it seems to be (see above). On way two consider the gradual pertinence of several disciplines to one theme, is to embrace the public perception. The combined appearance of theme and discipline in newspapers, journals and web databases provides an indication of pertinence. This approach is based on the assumption, that disciplines are more often mentioned together with a specific theme, if they are pertinent ones. The following approach uses a methodology transferred from the calibration paradigm (Nelson & Dunlosky, 1991): First, a measure for quality of the students discipline ratings was determined. Objective values were computed from the relative frequency of the co-occurrence of the theme ("tide and float") and each discipline ("e.g. physics") within Lexis Nexis, a large text data base. By reason, that the co-occurrence is just a frequency's subset of the theme and of the discipline, the value was divided by the sum of the overall frequencies. The strength of association between theme and each discipline has been used as an indicator of "objective" pertinence of the discipline to the topic. (Note that it is only 'objective' in the sense that it is reflected in the thousands of documents which make up the data base, and that we use it here as the norm for a comparison with of our students' assessments). The systematic relationship between student's discipline ratings and the objective value was determined by computing within-subject Goodman-Kruskal Gamma correlations ($G$). These were subsequently Z-transformed into indices. Median of indices overall students was $G = 0.58$ ($t(518) = 16.21$, $p < .001$).

The results can also be considered in a descriptive manner. Student's ratings differed between disciplines (Figure 1). An empirical trichotomy of the data gets obvious. Seven disciplines (oceanography, physics, astronomy, geo science, geology, life sciences, biology) with a mean rating above $M = 4.0$ are highly associated with the theme. The pertinence of five disciplines (meteorology, math, landscape ecology, hydrology, chemistry) is much more graded ($3.0 <$ mean rating $M < 4.0$). All other disciplines are mostly declined (mean rating $M < 2.0$). To sum up, student's ratings correlate significantly with the associations of disciplines and theme in a text data base. Furthermore, students clearly identify disciplines which have to contribute, maybe should contribute and definitely should not contribute.

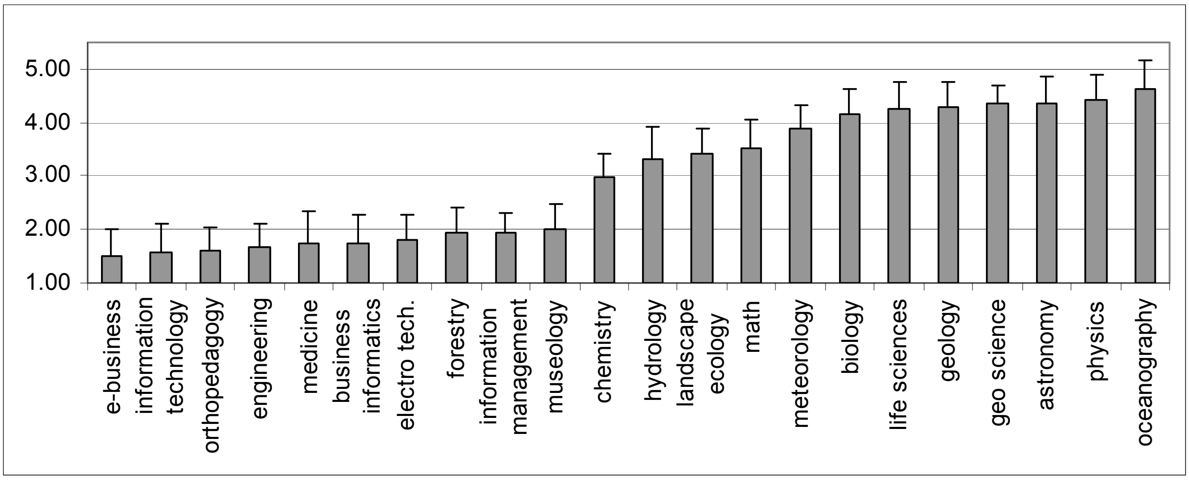



Figure 1. This figure visualizes the mean discipline ratings.

## Are discipline ratings impacted by epistemological beliefs and knowledge self-assessment?

In order to answer this question the experimental groups (straightforward versus sophisticated instruction text) as well as gender were included as dichotomous factors in a measurement model (linear regression). To account for the trichotomy in discipline ratings three dependent variables have been computed: "Mean rating of pertinent disciplines (RD)" (*N* = 7, *M* = 4.35, *SD* = 0.49), "Mean rating of disciplines with graded pertinence (GD)" (*N* = 5, *M* = 3.42, *SD* = 0.63) and "Mean rating of non-pertinent disciplines (ND)" (*N* = 10, *M* = 1.76, *SD* = 0.55). Interactions between errors of these three variables (RD, GD and ND) are accepted in the model. Furthermore, the knowledge self-assessment scale (KSA) was included in the measurement model. Fit indices indicate a sufficient model fit (Byrne, 2001). Relative chi-square ($\chi 2/df = 1.57$) and the comparative fit index ($CFI = 0.98$) exhibit an acceptable fit between the hypothetical model and the sample data as well as the point estimate of the root mean square error of approximation ($RMSEA = 0.03$). No significant interactions between independent variables (experimental group and gender) have been found. As shown in Figure 2, introduction text ($r = -.42$, $p < 0.001$) and gender ($r = -.29$, $p = 0.003$) significantly impacted knowledge self-assessment (KSA). Students who read the sophisticated introduction and female students significantly stated to know less about "tide and float". The independent variable "introduction text" also impacted significantly the "Mean rating of pertinent disciplines (RD)" ($r = .27$, $p = 0.002$). Students in the group with the sophisticated introduction text agree significantly stronger to the pertinent disciplines (RD). By tendency, students who read the sophisticated introduction agree also stronger to the disciplines with graded pertinence (GD) ($r = .17$, $p = 0.058$) and decline non-pertinent disciplines (ND) more ($r = -.05$, $p$ = n.s.). No relation between knowledge self-assessment (KSA) and discipline rating as well as between gender and discipline rating has been found.

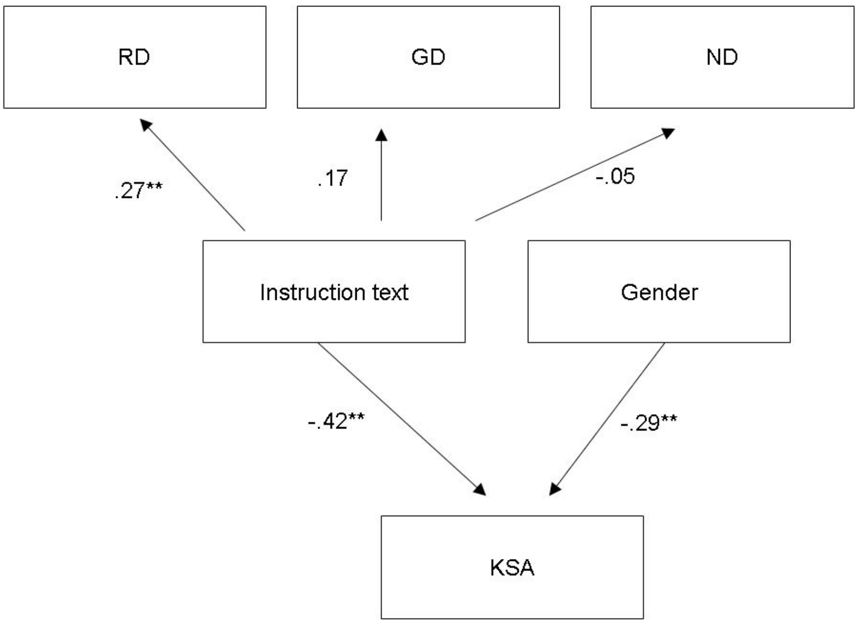



Figure 2. This figure visualizes the results of the measurement model ($N = 520$). (RD = Mean rating of pertinent disciplines, GD = Mean rating of disciplines with graded pertinence, ND = Mean rating of non-pertinent disciplines, KSA = knowledge self-assessment).

## Discussion

To rely on others gets increasingly important in modern society (Bromme, Kienhues & Porsch, 2009). A realistic view on own knowledge as well as knowledge about pertinence of disciplines are helpful requirements in academic and informal settings. Results of this study indicate that students are aware of discipline's pertinence and know which disciplines are pertinent to the theme "tide and float". They are able to assess disciplines' pertinence in a meaningful way. Student`s ratings are significantly correlated with findings in a large text data base. The pertinence ratings of the used disciplines to "tide and float", are concordant with the appearance of the combination of disciplines and theme in newspapers, journals and web databases. The descriptive consideration of the data reveals that next to pertinent and non-pertinent disciplines, some disciplines exhibit a graded pertinence to the theme. Empirically three groups of pertinence (pertinent disciplines, disciplines with graded pertinence and non-pertinent disciplines) could be established. The discipline pertinence ratings of these groups are partly explained by the epistemological sensitization. The results show that all students assume that there are some disciplines which are pertinent (oceanography, physics, astronomy, geo science, geology, life sciences, biology) or graded pertinent (meteorology, math, landscape ecology, hydrology, chemistry). But students who read the sophisticated introduction text agree stronger to the pertinent and graded pertinent disciplines. Students who read the straightforward introduction text preferred these disciplines less. This result indicates that students in the sophisticated sensitization group have more elaborated ideas about the pertinence of disciplines: They tend to include the pertinent and graded pertinent disciplines more when asked which discipline might be pertinent to a certain topic. The given topic is a quite multidisciplinary one. Students with strong ideas of the connectedness and variability of knowledge might tend to transfer their beliefs on ideas about the pertinence of disciplines. The results indicate that students are able to do this in a meaningful matter. Instead of assuming a general pertinence of all disciplines, students with more sophisticated beliefs made a better choice. The sophisticated epistemological sensitization induces a more differentiated choice between the disciplines. All disciplines which seemed, even only rudimentarily, pertinent were appraised higher.

The expected influence of the knowledge self-concept on disciplines' pertinence rating has not been found. Students concept about own knowledge about the topic doesn`t impact their ability to assign pertinent disciplines to the topic. We conclude (cautiously because it is a speculation based on the lack of empirical relationship) that ideas about who knows what (knowledge about the division of cognitive labor) are relatively independent from knowledge about the reference issue itself. Our cautious conclusion is in line with Keil et al, (2008) who reported awareness about the division of cognitive labor held by young children who clearly lack deep knowledge about the reference issue.

Gender influences the knowledge self-assessment independent from epistemological beliefs. Female students stated to know less. This result is in line with findings about gender specific self concept in academic contexts (e.g. Shaalvik & Rankin, 1990; Marsh & Craven, 1997; Jacobs et al., 2002). The theme "tide and float" is contextualized in school in the subjects as "physics" or "geography". Gender differences in self-concept in context of these "natural scientific" subjects are as expected. But these differences did not impact the rating of discipline's pertinence. Neither knowledge self-assessment nor gender itself has impacted student's choices.

The analyses uncovered further results. With regard to the content used here, the sophisticated epistemological sensitization causes students to asses more cautiously their own knowledge about the issue at stake. But this is not necessarily a misjudgment. Their knowledge self-assessment could be simply more realistic. Even if they have already superficial knowledge about the theme, most of the students have not dealt with the underlying mechanisms before. With regard to the research by Rosenblit & Keil (2002), we assume that the straightforward sensitization contributes to an "illusion of explanatory depth". In this study, the confrontation with epistemological uncertainties ("texture" and "variability") e.g. by detailing scientific controversies, might have helped the students to relativize their own knowledge about "tide and float" in a realistic way. Further studies should also take qualitative methods into account to provide an insight into why more sophisticated epistemological views make students more critical about their own knowledge.

There are, however, some limitations of this study: We did not capture underlying reasons for preferring a discipline but rather ask for student's spontaneous judgments. Additionally, we acknowledge that our epistemological sensitization might not have resulted in a sustained change of the students' epistemological beliefs. Within this study we focused on the relationship between epistemological views and pertinence assessments and the question, if the epistemological views which we have established within our experiment will be stable or not can be postponed to further studies. Nevertheless, our finding that it is possible to induce such an epistemological sensitization by a rather sparse intervention has interesting educational implications. The intervention which was used here for experimental reasons could be also used for educational aims. Of course, such sensitization might only work with specific topics. The implementation of the epistemological sensitization has to be tested further to use it in more real life settings (e.g. classrooms and textbooks) with more diverse populations. In everyday problem solving and decision making it is a quite common task to decide about disciplines pertinence in context of mostly unknown knowledge topics, but the capabilities which are necessary here have not been researched intensively so far. While our results are clearly confined to our sample (secondary school students) they do nevertheless demonstrate how such capabilities could be studied.

## References

Bergstrom, B., Moehlmann, B., & Boyer, P. (2006). Extending the testimony problem: evaluating the truth, scope, and source of cultural information. *Child Development, 77*, 531-538.

Bromme, R., Kienhues, D., & Porsch, T. (2009). Who knows what and who can we believe? Epistemological beliefs are beliefs about knowledge (mostly) attained from others. In L. D. Bendixen & F. C. Feucht (Eds.), *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice* (pp. 163-193). Cambridge: Cambridge University Press.

Buehl, M. M. & Alexander, P. A. (2001). Beliefs about academic knowledge. *Educational Psychology Review, 13*(4): 385-418.

Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of the Learning Sciences, 12*, 53-90.

Harris, P.L. (2001). Thinking about the unknown. *Trends in Cognitive Sciences, 5*, 494-498.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*(1), 88-140.

Hofer, B. K., & Pintrich, P. R. (Eds.). (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, NJ: Lawrence Erlbaum.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's selfcompetence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*, 509-527.

Keil, F. C., Stein, C., Webb, L., Billings, V. D., & Rozenbilt, L. (2008). Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science, 32,* 259-300.

Kienhues, D., Bromme, R., & Stahl, E. (2008). Changing epistemological beliefs: The unexpected impact of a short-term intervention. *British Journal of Educational Psychology*, *78*, 545-565.

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development, 73*(4), 1073-1084.

Marsh, H. W. & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal-external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56-67.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.),*Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131-198). Orlando, FL: Academic Press.

Nelson, T. O. & J. Dunlosky (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: the delayed-JOL effect. *Psychological Science, 2*(4): 267-270.

Pieschl, S., Bromme, R., Porsch, T. & Stahl, E. (2008). Epistemological sensitisation causes deeper elaboration during self-regulated learning. *International perspectives in the learning sciences: Cre8ting a learning world*. *Proceedings of the Eighth International Conference for the Learning Sciences* - ICLS 2008, Vol. 2 (pp. 2-213 - 2-220). London: Lulu Enterprises.

Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*, 521-562.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*(3), 498-504.

Shaalvik, E. M., & Rankin, R. J. (1990). Math, verbal, and general academic self-concept: The Internal/External Frame of Reference models and gender differences in self-concept structure. *Journal of Educational Psychology*, *82*, 546–554.

Stahl, E. & Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction, 17* (6), 773-785.

Stahl, E., Pieschl, S., & Bromme, R. (2006). Task complexity, epistemological beliefs and metacognitive calibration: An exploratory study. *Journal of Educational Computing Research*, *35* (4), 319-338.

Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative understanding of school subjects: Past, present, and future. *Review of Educational Research, 75*, 125-157.

Trautwein, U. & Lüdtke, O. (2007). Predicting global and topic-specific certainty beliefs: Domain-specificity and the role of the academic environment. *British Journal of Educational Psychology, 77*, 904-934.

Zeidler, D. L., Sadler, T. D., Applebaum, S., & Callahan, B. E. (2009). Advancing reflective judgment through socioscientific issues. *Journal of Research in Science Teaching*, *46*, 74-101.

# Free, open, online, mathematics help forums: The good, the bad, and the ugly

Carla van de Sande, Arizona State University, carla.vandesande@asu.edu

**Abstract:** Free, open, online, help forums link students with volunteer helpers who have the time, knowledge, and willingness to provide assistance with specific problems from coursework. Some forums endorse a select learning theory, whereas others do not espouse any one approach. This study shows how student activity differs in a forum that supports Cognitive Load Theory (CLT) and one that does not embrace the practice of any single learning theory. A sample of 100 exchanges on the calculus concept of limit were collected from the archives of a forum of each type and compared for student contributions to the construction of the solution and student initiative in expressing resolution. The results show that the forum supporting CLT evidences lower levels of student activity (both initially and following helper intervention), and favors weak over strong expressions of resolution. The designation of the good, bad, and ugly depends on the reader's epistemological stance.

Today's students are using the Internet as a resource for completing their assignments through participation in open, online help forums. Many such forums, covering a wide variety of school subject areas, are found on websites that are accessible to the general public (open) and allow students everywhere to communicate anonymously and asynchronously with volunteers around the world who have the time, willingness, and experience to help them. Students post queries (usually problem-specific questions from assignments) on these forums when they are seeking help constructing a solution to a problem or when they are seeking verification of a solution that has been constructed. In order to help students construct solutions to their problems, some forums allow any member to respond to the student (e.g., as a helper) and participate in ongoing exchanges. Help is this case is a spontaneous activity, and we refer to these sites as Spontaneous Online Help (SOH). In short, these forums afford convenient, accessible, efficient help to thousands of students, while, at the same time, they transform help seeking and helping from a private, one-on-one activity bounded by physical walls into a public, communal endeavor that spans geo-political boundaries (2007). This contribution to the democratization of education is part of "the good" (Larreamendy-Joerns & Leinhardt, 2006); what constitutes "the bad" and "the ugly" depends on the epistemological stance of the reader.

These forums exist primarily for the purpose of helping students. How do they then tackle the monumental and ill-defined task of defining *what it means to help* within their community, especially given the diversity of the members? The answer is that some forums explicitly endorse a select learning theory, whereas others do not espouse any one approach. The purpose of this paper is to explore the nature of student activity in a mathematics forum that uses Cognitive Load Theory (CLT) as its modus operandi (Cramster) versus another forum that does not advocate any particular theory of learning (FreeMathHelp). In both cases, the community of helpers is not required to adhere to any mandated pedagogical practices, but rather the corps of helpers has self-selected a forum in which to volunteer. Regardless of enforcement, however, the position taken by the forum administration and helpers on theories of learning and instruction is expected to show up in the manner in which students use the forum as they seek help completing assignments.

## Conceptual Framework

### Help seeking

These problem statements that students post on the help forums generally stem from routine exercises, commonly referred to as closed tasks. However, if a student is unable to solve such a task without assistance, then the "exercise" has become a "problem" from the student's perspective (Selden, Selden, Hauk, & Mason, 2000). In this way, the forums are instantiations of help seeking, an activity that no longer carries the stigma of incompetence or stupidity (Nelson-Le Gall, 1981). In fact, quite the opposite may be true, as students engaged in help seeking can be proactively taking responsibility for acquiring new knowledge and skills (Resnick & Nelson-Le Gall, 1997, p. 150). Help seeking, *with appropriate performance goals*, has strategic value in the learning process. Nelson-Le Gall (1985) distinguished between "executive" or dependency-oriented help seeking that is associated with performance goals ("Just tell me the answer") and "instrumental" or mastery-oriented help seeking ("Give me a clue!"). Now, given that students are seeking help using open, online, homework help forums to construct solutions to routine exercises, the question is: In what ways are students acquiring the knowledge and skills that allows them to complete their assignments?

Alternative learning theories advocate different definitions of what type of help is conducive to learning. I now turn to one learning theory that has a lot to say about how students learn.

## Cognitive Load Theory (CLT)

CLT is a theory of learning developed by Sweller (1988) that is based on the limited capacity of working memory and emphasizes the need for instructional design to minimize cognitive processing or load. Determining pedagogical practices that are consistent with this theory has therefore been a central part of the research agenda (Kirschner, 2002). The use of worked examples is one such practice that has been extensively researched (Atkinson, Derry, Renkl, & Wortham, 2000). The provision of worked examples as a study tool has been shown to facilitate near transfer, require less acquisition and performance time, and require less mental effort than engaging in problem solving or exploration (Paas & Van Merriënboer, 1993; Sweller & Cooper, 1985; Tuovinen & Sweller, 1999; Van Merriënboer, Schuurman, De Croock, & Paas, 2002), at least for students who are really unfamiliar with the material (Kalyuga, Chandler, Tuovinen, & Sweller, 2001).

Of course, no self-respecting cognitive load theorist would claim that worked examples are, in and of themselves, a magic bullet. It is the way in which students engage with worked examples that determines how much help they provide. For instance, students who self-explain worked examples learn more deeply than students who simply read through them (Chi, Bassok, Lewis, Riemann, & Glaser, 1989) or copy them (Chi, 2009).

The Cramster forum advocates providing worked examples as an effective means of providing help. Drawing explicitly on CLT, they claim that this is the best and most efficient way of helping students with their homework assignments.

## Methods

The objective of this project was to investigate the nature of student activity on sites with differing adherence to theories of learning. Two open, online, mathematics help forums were chosen, one that espouses CLT and one that does not endorse any given learning theory. In addition, Cramster.com has a reputation system (Dellarocas, 2003) in place that allows students to rate the contributions of helpers, whereas FreeMathHelp does not. Associated with this distinction is the fact that Cramster offers subscription levels that increase the weight of ratings for others' contributions.

Because of the exploratory nature of this work, an observational methodology was adopted (Goodyear, Jones, Asensio, Hodgson, & Steeples, 2005). In order to draw comparisons across the two sites, I gathered 100 exchanges from the archives of each forum on a single calculus concept, the limit, by conducting a search on "lim*." Although there are arguably many concepts in calculus instruction that surface on the help forums and that could have served my purpose, the limit concept took precedence, being both foundational and poorly understood by students (Cornu & Tall, 1991; Szydlik, 2000; Tall & Vinner, 1981). The Cramster postings were first gathered in December, 2008 and the search was performed on exchanges dating back from November, 2008. These dated back to September, 2008 and were compared with a set of analogous exchanges from the archives of FreeMathHelp that had been gathered for previous research. The FreeMathHelp corpus dated from April, 2008 to January, 2007. The differing time spans covered by the searches (3 months for Cramster versus 15 months for FreeMathHelp) reflects the relative popularity of the two sites; Cramster receives much more traffic than FreeMathHelp. Because Cramster offers subscription and FreeMathHelp operates strictly gratis, the Cramster search specified that only exchanges initiated by nonpaying Cramster members be included. However, the Cramster sample may include some students whose status changed between the time of posting and the time of data collection.

## Site Descriptions

### Cramster

www.cramster.com (Cramster) is a global study community for help in many subjects, including physics, math, science, and engineering. The site was launched in 2003 by Cramster, Inc., a private company headquartered in Pasadena, California. The site advocates CLT and has been recently endorsed by well-known educational theorist, John Sweller, who is also a member of the Cramster Academic Advisory Board: "Cramster provides an effective learning environment for difficult and complex concepts. I strongly recommend to any student."

In addition to the discussion forum or Q&A board, the site provides access to study materials such as textbook solutions, topic notes, sample problems, and practice exams that were created by the administration, indexed from the web, or contributed by members. Although general membership is free, participants can elect to subscribe monthly ($9.95/month) or annually ($49.95/year) to enjoy additional site benefits and access. Members have access to user profiles that include self-volunteered information on school, major, and expertise. In each contribution, members are characterized by a self-selected username, an optional avatar (static or animated picture that represents the user), board level (ranging from Rookie to Oracle), and "karma."

Karma is acquired through participation in site activities (such as answering queries on the Q&A board, submitting solutions to textbook problems, challenging others' solutions, providing notes, quizzes, outlines, etc., and referring friends to become members) and is based on an intricate point system. For the Q&A board that is the focus of this study, karma points depend on the rating awarded by the member who asked the question (Lifesaver, Helpful, Somewhat Helpful and Not Helpful), their membership status, and the difficulty level of the board. Karma points can be redeemed for gift certificates and prizes. On the part of students, rating responses is voluntary but does affect the student's "respect score" (the number of answers the student has rated divided by the number of answers provided to that student).

In addition to standard Terms of Use for activity on the Internet, Cramster has an Anti-Cheating policy posted: "Copying solutions or posting unexplained final answers on the Q&A Board promotes completion without comprehension, and that's something we don't support on this site."

### FreeMathHelp

www.FreeMathHelp.com (FreeMathHelp) is an advertisement-supported mathematics help portal established in 2002 by Ted Wilcox, who was, at the time, an enterprising high school junior. The site contains 10 homework help forums organized by subject area (ranging from arithmetic and pre-algebra to calculus and differential equations). The sole requirement for becoming a forum member is registration (which entails agreeing to abide by terms for permissible content and/or conduct, providing a username and e-mail address, and selecting a password). Forum members can initiate threads in a discussion forum (e.g., as students posting mathematics questions) and can respond to others' posts (e.g., as helpers). Forum members also have access to user profiles that include self-volunteered information on occupation, residence, contact information, as well as statistics on discussion board activity.

This site does not espouse any one pedagogical approach. At one time, a lengthy and passionate debate was held in the community forum by prominent helpers concerning the nature of the assistance that helpers should be allowed to provide. The suggestions covered a wide range of possibilities, from banning helpers who provided "copy-ready" solutions to creating a separate forum for those who wished to provide worked solutions. The final decision, made by Ted as forum administrator, was to allow helpers to practice any form of pedagogical assistance, with a request for helpers to respect and not interfere with the helping strategies employed by others in the community.

Forum "netiquette" (Shea, 1994) is located in a "sticky" that is the lead posting within each help forum. Although the forum does not ascribe to any particular learning theory, students are encouraged to show all of their work. The rationale is that, by doing so, helpers are more able to detect the source of student difficulty and the forum can operate efficiently. If, for instance, no work is shown, then helpers may assume that the student only needs help getting started, or, if only partial work is shown, then helpers will not be privy to errors in undisclosed portions of the solution.

### Analyses

The objective was to illuminate differences in student behavior depending on the forum. Analyses of student activity were designed to capture the trajectory of understanding from the perspective of the student over the course of a given exchange: initial and post-intervention (during), and final. Each exchange was therefore examined for the presence of student assertions and proposals of mathematical actions, both in the *initial* posting in which a query was presented as well as following helper intervention *during* the ensuing thread. Examples of actions that were taken as evidence of student activity included suggestions of a solution procedure, proposals of solution steps, and follow-throughs on others' proposals. In order to be inclusive in this analysis of help-seeking activity, even instances that were hedged or offered in a hesitant manner, were counted. The last contribution made by the student in each exchange and the ratings provided by students on the Cramster forum were used as indicators of the *final* state of understanding, that is, as an expression of how the student felt about the resolution of the problem. The strength of resolution was further characterized as "weak" (expressions of gratitude) or "strong" (reflection on why or how interaction was helpful). These overt activities (contributing to the construction of the solution and initiative for expressing resolution) reflect student agency (Greeno, 2006) and characterize interactive activities (Chi, 2009).

## Results

### Contributing to the construction of solution

Figure 1a shows the percentage of threads in each forum according to the presence of student suggestions and proposals for action in the initial posting and subsequent thread: [0 0] (no student activity), [1 0] (student activity only in the initial posting), [0 1] (student activity only after the initial posting), and [1 1] (student activity in both initial and subsequent postings in the thread).

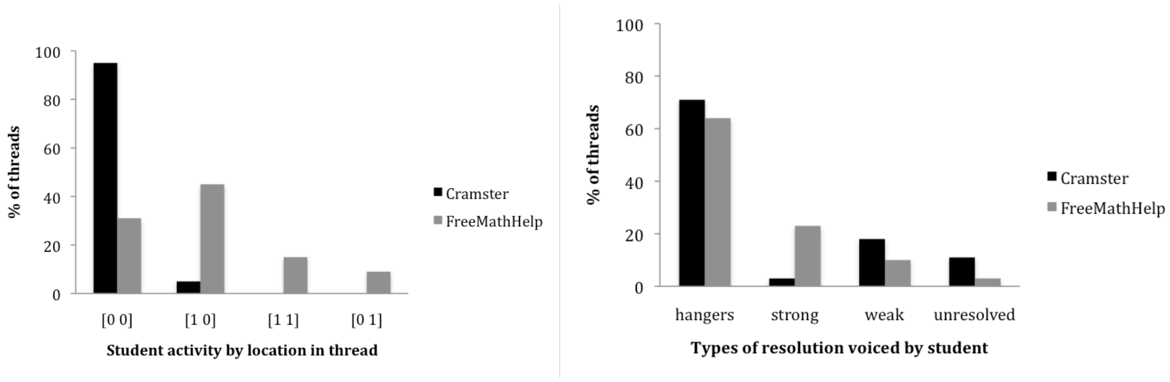



Figure 1. a) Student contributions to construction of solution; b) Student expressions of resolution by type

There are two noteworthy results here: first, Cramster contained *no* threads in which student activity increased following helper intervention ([0 1]) or in which student activity was present throughout the exchange ([1 1]), and second, *virtually all* (95%) of this site's exchanges were completely devoid of any student mathematical activity ([0 0]). On this site, students pose their queries without showing work and the helpers almost exclusively contribute worked solutions (see Figure 2). In contrast, the majority of the exchanges (69%) in FreeMathHelp contained evidence that students were actively contributing to the construction of the solution, either in the initial post or in response to helper assistance (see Figure 3).

The most sensible explanation for these results can be found in the differential nature of the help provided on the two forums. On Cramster, 98% of the helpers' first interventions (the first response received by the student) consisted of a full worked solution; not surprisngly, there was no need then for students to contribute to the construction of the solution. On FreeMathHelp, however, only 23% of the first interventions could be classified as a worked solution. Thus, in many cases, a worked solution was not provided by a helper, and students could still contribute to the construction, if they chose to do so.

## Initiative for expressing resolution

There are several ways that a participant can indicate that an issue has (or has not) been resolved. First of all, participants can be silent and opt not to further contribute to an exchange. Silence in computer-mediated exchanges may indicate acceptance or rejection of another's contributions and does not offer evidence for (or against) the achievement of resolution. Thus, in the forum discussions, if a student does not return to the exchange following helper interventions, it is not clear whether the student feels that the issue has been settled or not. I refer to exchanges of this type as "*hangers*" since other forum participants are, in some sense, left hanging regarding the helpfulness of their contributions. On the other hand, when a student does explicitly acknowledge helpers' contributions, they can do so in either a *weak* or *strong* manner. For instance, an expression of appreciation, such as "Thank you," indicates a weak level of resolution on the part of the student since this may simply be a residual of polite manners, that is, a customary response to receiving assistance. Also, if a student simply rates the quality of received help, then this rating is indicative of weak resolution since it does not signify how the interaction was helpful. In contrast, the contribution of mathematical actions (e.g., the presentation of a solution to the problem) and assessments (e.g., reflections on the ways in which the intervention helped) are stronger indications that the issue has been resolved to the satisfaction of the student. Finally, an exchange can evince a lack of resolution or be *unresolved*, when a student receives no response to a query or receives a refusal from forum tutors to provide further assistance.

Figure 1b shows the percentage of exchanges for each forum in which resolution could not be determined (hangers), in which resolution was evident and the strength of the expression (weak versus strong), and in which the issue was unresolved.

Given the informal and transitory nature of these interactions, it is not surprising that the majority of exchanges were characterized as hangers in both forums. However, those that did evince some degree of resolution showed a different pattern in Cramster versus FreeMathHelp. First of all, in Cramster 11% of the exchanges were unresolved, with 4% demonstratively unresolved following helper intervention. In contrast, in FreeMathHelp, only 3% of the exchanges were unresolved, which is the worst outcome of forum participation from a student's perspective. Also, although the expression of resolution is promoted in Cramster because of the "respect score," there was a higher percentage of expressions of resolution on FreeMathHelp (33% versus 21%), as seen by collapsing over resolution strength. A distinction between strong and weak resolution showed an additional difference in this aspect of student activity across the two forums. In Cramster, expressions of weak resolution (18%) were more prevalent than strong (3%), whereas this trend was reversed in FreeMathHelp, in which strong resolution was indicated in 23% of the exchanges, whereas 10% showed weak resolution. It appears that Cramster favors the demonstration of weak resolution (if any), in which students either provide a rating or express thankfulness, but fall short of remarking on the way in which they received useful help and gained understanding (see Figure 2). In contrast, something about the interaction in FreeMathHelp encourages students to express resolution, with no incentive, reflect publicly on the way that the help received contributed to their understanding of the mathematics in question (see Figure 3).

## Examples

### Cramster

Figure 2 shows a prototypical example of a Cramster exchange derived from the in-depth analyses of student activity. The student, **Ali**, did not contribute in any way to the construction of the solution ([0 0]) and gave the helper, **fatbuddha**, a lifesaver rating (weak resolution) following the provision of a (sparsely) annotated worked solution. **Ali** typed in the text of the exercise, **fatbuddha** constructed the solution, and **Ali** evaluated the reply. The entire transaction took only 11 minutes, where the median length of time until a student received a first reply in the Cramster sample of 100 exchanges was 26 minutes. The exchange is remarkably impersonal and devoid of social interaction.

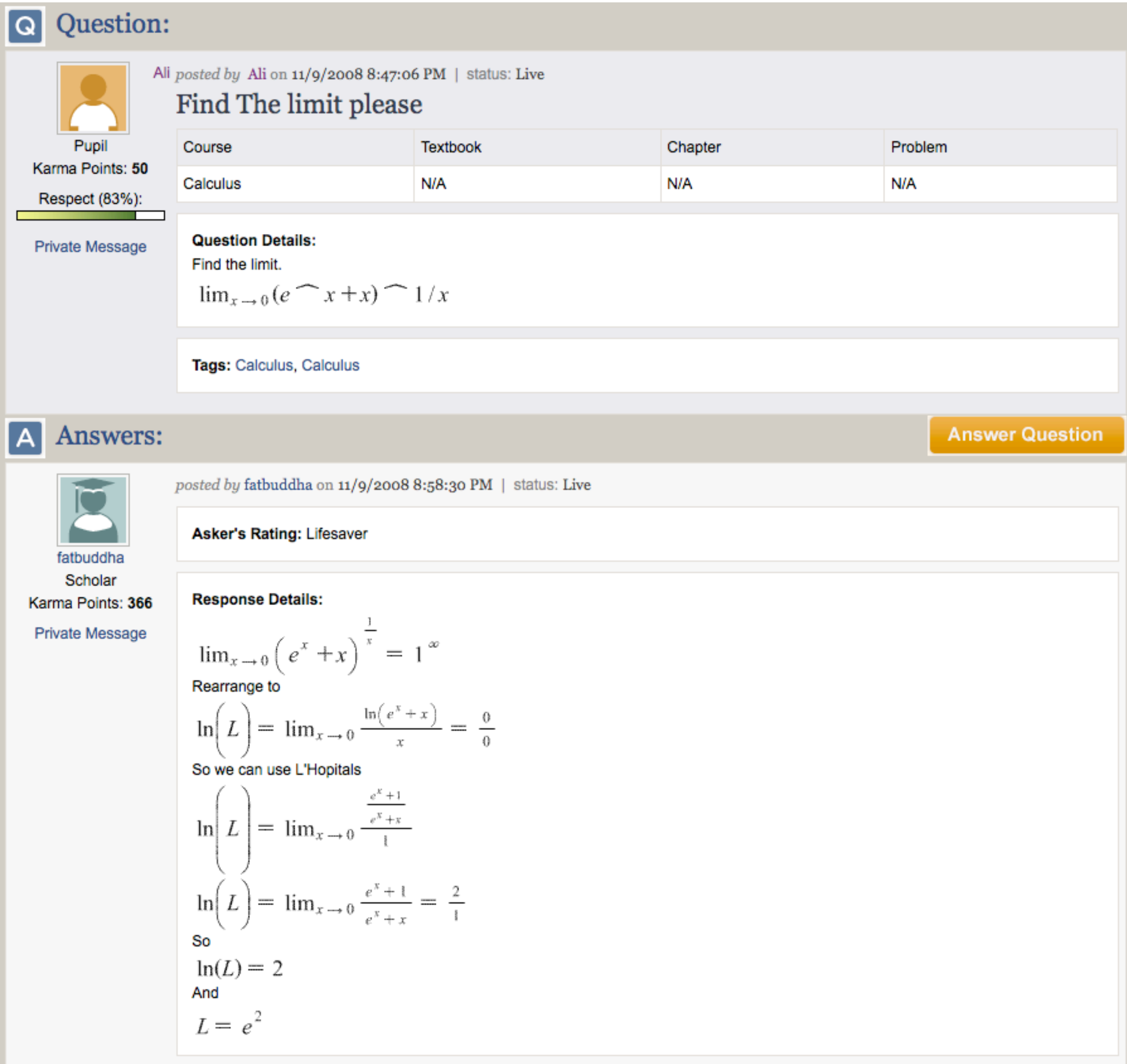
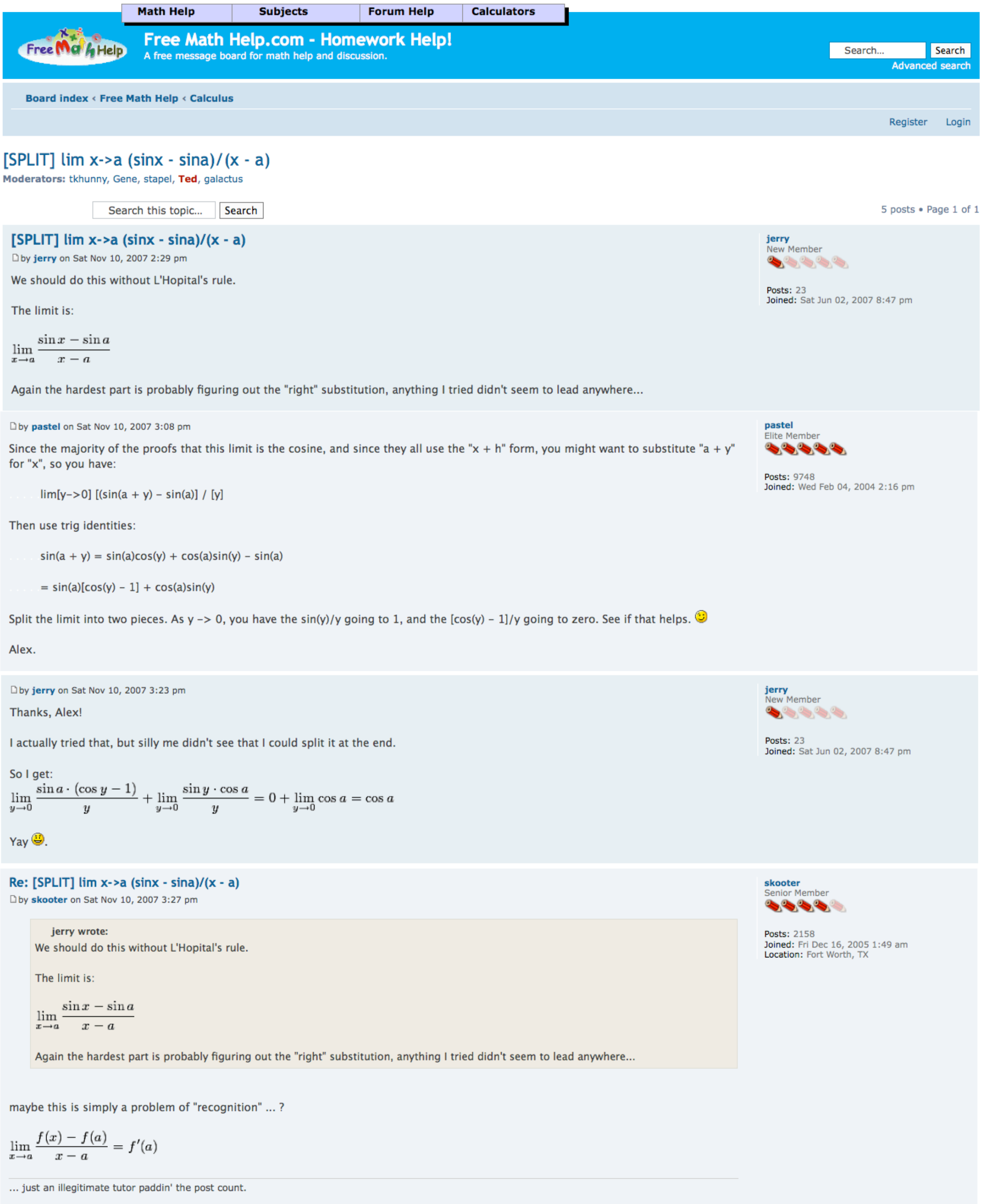



Figure 2. Cramster exchange showing absence of student activity and weak resolution.

### FreeMathHelp

Figure 3 contains an example of an exchange from FreeMathHelp. Although not prototypical in the sense that the student, **jerry**, contributes to the construction of the solution throughout the exchange ([1 1]), this example captures the flavor of many interactions that take place on this help forum. When **jerry** poses the query, he includes a suggestion, albeit hedged, for constructing the solution, namely substitution. **Jerry** suggests that the goal to the solution is finding the right substitution, but does not specify the substitutions that he has tried but which "didn't seem to lead anywhere." The first helper, **pastel**, provides the substitution, requisite trigonometric identities and outlines the rest of the solution, but stops shy of providing a worked solution. Shortly after this help is received, **jerry**, returns to the exchange, comments on the insight he has gained because **pastel** showed him the rearrangement of terms, and contributes the rest of the solution. Although **jerry** has indicated that the issue has been resolved from his perspective, a second helper, **skooter**, comes on the scene and suggests through hinting that the limit can also be framed alternatively as the definition of a derivative. In response, **jerry**, returns once again, expresses appreciation, and comments on how he has gained new insight on this problem, thanks to **skooter**. The exchange gives a definite sense that **jerry** has experienced "Aha!" moment(s) and an almost tangible feeling of excitement that is expressed through **jerry's** use of [very happy] emoticons.



Math Help | Subjects | Forum Help | Calculators

Free Math Help.com - Homework Help!
A free message board for math help and discussion.

Search... Search
Advanced search

Board index ‹ Free Math Help ‹ Calculus

Register Login

[SPLIT] lim x->a (sinx - sina)/(x - a)
Moderators: tkhunny, Gene, stapel, Ted, galactus

Search this topic... Search

5 posts • Page 1 of 1

[SPLIT] lim x->a (sinx - sina)/(x - a)
by jerry on Sat Nov 10, 2007 2:29 pm

We should do this without L'Hopital's rule.

The limit is:

$$\lim_{x \to a} \frac{\sin x - \sin a}{x - a}$$

Again the hardest part is probably figuring out the "right" substitution, anything I tried didn't seem to lead anywhere...

jerry
New Member
Posts: 23
Joined: Sat Jun 02, 2007 8:47 pm

by pastel on Sat Nov 10, 2007 3:08 pm

Since the majority of the proofs that this limit is the cosine, and since they all use the "x + h" form, you might want to substitute "a + y" for "x", so you have:

lim[y->0] [(sin(a + y) - sin(a)] / [y]

Then use trig identities:

sin(a + y) = sin(a)cos(y) + cos(a)sin(y) - sin(a)

= sin(a)[cos(y) - 1] + cos(a)sin(y)

Split the limit into two pieces. As y -> 0, you have the sin(y)/y going to 1, and the [cos(y) - 1]/y going to zero. See if that helps.

Alex.

pastel
Elite Member
Posts: 9748
Joined: Wed Feb 04, 2004 2:16 pm

by jerry on Sat Nov 10, 2007 3:23 pm

Thanks, Alex!

I actually tried that, but silly me didn't see that I could split it at the end.

So I get:

$$\lim_{y \to 0} \frac{\sin a \cdot (\cos y - 1)}{y} + \lim_{y \to 0} \frac{\sin y \cdot \cos a}{y} = 0 + \lim_{y \to 0} \cos a = \cos a$$

Yay.

jerry
New Member
Posts: 23
Joined: Sat Jun 02, 2007 8:47 pm

Re: [SPLIT] lim x->a (sinx - sina)/(x - a)
by skooter on Sat Nov 10, 2007 3:27 pm

> jerry wrote:
> We should do this without L'Hopital's rule.
>
> The limit is:
>
> $$\lim_{x \to a} \frac{\sin x - \sin a}{x - a}$$
>
> Again the hardest part is probably figuring out the "right" substitution, anything I tried didn't seem to lead anywhere...

maybe this is simply a problem of "recognition" ... ?

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = f'(a)$$

... just an illegitimate tutor paddin' the post count.

skooter
Senior Member
Posts: 2158
Joined: Fri Dec 16, 2005 1:49 am
Location: Fort Worth, TX

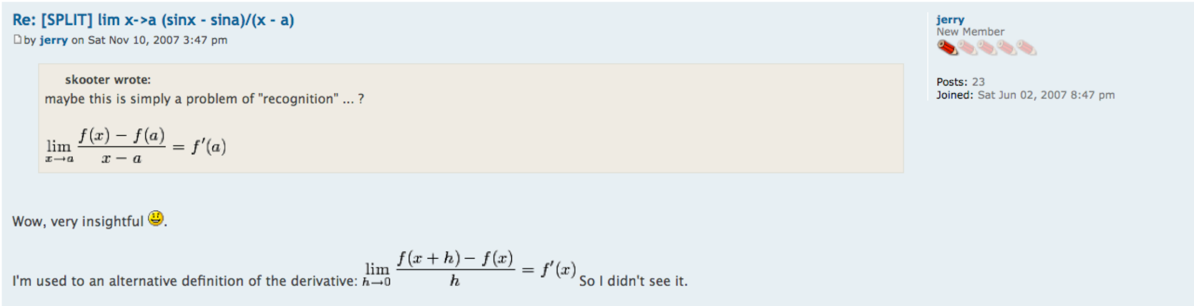



Figure 3. FreeMathHelp exchange showing student activity and strong resolution.

## Discussion

The message of this paper is that student activity looks dramatically different on two open, online, calculus, help forums. On Cramster, the forum in which CLT is the modus operandi, students do not contribute to the construction of the solutions, and favor weak over strong expressions of resolution. A typical dialogue frame (Graesser, Person, & Magliano, 1995) for interaction on Cramster consists of a student presenting a problem statement, a helper contributing a worked solution, and, perhaps, the student rating the help received. In fact, the primary instructional sequence is the traditional IR/E pattern from classroom discourse (Mehan, 1982), but with an monumental twist – the role of instructor and student is reversed. It is the student who initiates the interaction by asking a question (I), the helper who replies with the answer (R), and the student who evaluates the response (E). In contrast, on FreeMathHelp, the forum in which helpers to not adhere to any particular learning theory, students often initiate or follow-up on the construction of the solutions and favor strong over weak resolution by sharing how the interaction was helpful. Here, there is evidence of extended dialogue frames that include student contributions and reflections. Evidence for computer-supported collaborative learning (such as the joint construction of solutions) can be seen in interactions in FreeMathHelp.

However, the observational methodology employed by this project prohibits making stronger claims than the existence of these differences. First, only published contributions were part of the analysis. Private messaging is a feature of both forums, and may be more frequently used in Cramster since helpers can only receive karma for their initial response to a student. This policy may well encourage students and helpers to communicate off-line. Anonymity adds another layer of complexity to interpreting results, since participants could have multiple usernames linked to different accounts. Fortunately, however, multiple participants cannot have the same username. Finally and more importantly, the measures cannot speak to the effect of forum participation on student performance in the classroom. The dramatic difference in the nature of student activity across the two forums raises the question whether student learning also dramatically differs. This is a question that must be answered in the context of an experimental study.

Open, online, help forums are being used by students throughout the world and have become part of their learning experience. As an emergent resource, we must determine their potential for instruction and learning. In particular, there is a pressing need to delve into what constitutes the good, the bad, and the ugly for this resource that students and helpers are frequenting. Returning to the title of this paper, what can we say at this point?

- The Good: Open, online forums are a means of connecting students seeking help with those who can provide it in an efficient, cost-effective manner. Students who might not otherwise have access to help are able to ask for help in a non-threatening environment. The forums also present an opportunity for educators and researchers to reach students beyond the walls of the classroom.
- The Bad: What counts as "bad" depends on the epistemological stance of the reader. Different forums offer different definitions for what it means to help students. From the perspective of a cognitive load theorist, providing worked examples is not bad. However, a constructivist would rather have students construct their own solutions and would probably consider Cramster to be an anathema.
- The Ugly: Given the open nature of the forums, official forum policies and educational philosophies can be abused or ignored. Things can get "ugly" in any forum, regardless of its ideals. Students can abuse the forums by asking for too much help, by mindlessly copying solutions, and by trying to use the forum to replace instruction.

John Sweller's recent endorsement of Cramster is evidence that education researchers are becoming aware of (even associated with) open, online, homework help forums. This paper intends to start a discussion on the design, use, and analysis of these learning activities that are happening "in the wild." One of the most important questions that this exploration has raised but admittedly danced around is whether worked solutions are an effective choice for helping students in the forums, and, if they are, then why forums, such as FreeMathHelp, that offer alternative approaches thrive. I will say that Cramster paints a picture of mathematics as an activity in which the goal is to reach answers, and the faster this is accomplished, the better. Even if students are using the solutions they get on Cramster as worked examples to help them complete additional exercises on their own, these students are still being sent a message about the nature of mathematics that seems problematic for those who love the subject and want students to share this passion.

## References


Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181-214.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities *Topics in Cognitive Science, 1*, 73-105.

Chi, M. T., Bassok, M., Lewis, M. W., Riemann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning how to solve problems. *Cognitive Science, 13*, 145-182.

Cornu, B., & Tall, D. (1991). Limits *Advanced Mathematical Teaching.* (pp. 153-166). Boston: Kluwer.

Dellarocas, C. (2003). The digitization of word of mouth: Promises and challenges of online feedback mechanisms. *Management Science, 49*(10), 1407-1424.

Goodyear, P., Jones, C., Asensio, M., Hodgson, V., & Steeples, C. (2005). Networked Learning in Higher Education: Students' Expectations and Experiences. *Higher Education, 50*(3), 473-508.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*, 495-522.

Greeno, J. G. (2006). Learning in activity. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 79-96). New York: Cambridge University Press.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Applied Educational Psychology, 93*, 579-588.

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1-10.

Larreamendy-Joerns, J., & Leinhardt, G. (2006). Going the distance with online education. *Review of Educational Research, 76*(4), 567-605.

Mehan, H. (1982). The structure of classroom events and their consequences for student performance. In P. Gilmore & A. A. Glatthorn (Eds.), *Children In and Out of School* (pp. 59-87). Washington, D.C.: Center for Applied Linguistics.

Nelson-Le Gall, S. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review, 1*, 224-246.

Nelson-Le Gall, S. (1985). Help-seeking behavior in learning. *Review of Research in Education, 12*, 55-90.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: an approach to combine mental effort and performance measures. *Human Factors, 35*, 737-743.

Resnick, L. B., & Nelson-Le Gall, S. (1997). Socializing intelligence. In L. Smith, J. Dockrell & P. Tomlinson (Eds.), *Piaget, Vygotsky and beyond* (pp. 145-148). London: Routledge.

Selden, A., Selden, J., Hauk, S., & Mason, A. (2000). Why can't calculus students access their knowledge to solve non-routine problems? In E. Dubinsky, A. Schoenfeld & J. Kaput (Eds.), *Research in Collegiate Mathematics Education* (pp. 128-153): American Mathematical Society.

Shea, V. (1994). *Netiquette*. San Francisco: Albion.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Science, 12*, 257-285.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59-89.

Szydlik, J. E. (2000). Mathematical beliefs and conceptual understanding of the limit of a function. *Journal of Research in Mathematics Education, 31*, 258-276.

Tall, D., & Vinner, S. (1981). Concept Image and Concept Definition in Mathematics with Particular Reference to Limits and Continuity. *Educational Studies in Mathematics, 12*(2), 151-169.

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334-341.

van de Sande, C., & Leinhardt, G. (2007). Online tutoring in the Calculus: Beyond the limit of the limit. *Éducation et Didactique, 1*(2), 115-154.

Van Merriënboer, J. J. G., Schuurman, J. G., De Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: effects on cognitive load, transfer test performance and efficiency. *Learning and Instruction, 12*, 11-37.

# Activity-Theoretical Research on Science Teachers' Expertise and Learning

Cory T. Forbes, University of Iowa, N252 Lindquist Center, Iowa City, IA, 52242, cory-forbes@uiowa.edu
Cheryl Ann Madeira, University of Toronto, 252 Bloor St. W., Toronto, ON, M5S 2V6, cmadeira@oise.utoronto.ca
James D. Slotta, University of Toronto, 252 Bloor St. W., Toronto, ON, M5S 2V6, jslotta@oise.utoronto.ca

**Abstract:** Teachers serve as critical mediators of student learning. As such, teachers' expertise and learning remain important foci for theoretical development and empirical research. Cultural-historical activity theory (CHAT) has been forwarded as an underutilized but potentially powerful tool for educational research, including teachers' expertise, practice, and learning. However, as yet, little CHAT-based research has been undertaken focused on science teachers and teaching. In this paper, we draw upon two such empirical studies in which CHAT was used as an explicit theoretical and analytical framework to explore CHAT-based perspectives on science teacher learning. We present findings from these studies to highlight important themes in CHAT-based research on science teachers' learning in and from practice. These findings can not only inform programmatic efforts to better promote teachers' learning, but also theoretical perspectives on science teachers' expertise, practice, and learning across the science education and learning sciences communities.

## Background and Theoretical Framework

In this paper, we present results from two CHAT-based studies of science teacher learning and practice to discuss theoretical contributions of CHAT to perspectives on science teacher learning and practice. In Study A, Forbes uses CHAT to investigate preservice elementary teachers' curriculum design and development of pedagogical design capacity for inquiry during the final year of their formal teacher education. In Study B, Madeira and Slotta use CHAT to investigate how lesson planning, enactment, reflection and peer exchange influence experienced secondary science teachers' development of pedagogical content knowledge (PCK) in a professional development context. Full findings from each of these studies have been presented previously. We therefore summarize here the study populations and designs, their explicit theoretical and analytical grounding in CHAT, the CHAT-based models we employed, methods of data collection and analyses, and main findings from each study. The primary purpose of this paper is to instead synthesize findings across these two studies to highlight and explore implications they hold for CHAT-based perspectives on teachers' learning and practice. These new theoretical insights can inform future research on science teachers and teaching.

### Situated Perspectives on Teachers' Knowledge and Expertise

Teaching is a complex and dynamic activity that requires teachers to develop robust expertise in order to best promote and facilitate student learning of science through inquiry. Teachers' expertise, and the manner in which it develops over time (i.e., learning), is fundamentally situated and embedded in contexts of practice. Participation in these communities of practice (Wenger, 1998), activity systems (Engeström, 1987), or ecosocial systems (Lemke, 2000) is one characterized by the process of semiosis, or meaning-making. Regular patterns of semiotic activity within such contexts, which develop over time and therefore have histories of their own, are characterized as practices. Through participation in existing practices, participants become more established participants and practitioners by aligning their own skill sets to conditions afforded in the setting where the activity occurs (Barab & Roth, 2006). However, by working to reconfigure constituent elements of existing practices, individuals can also fundamentally alter them. Learning, then, is defined as engaging in fundamentally new forms of practice (Engeström, 1987; Lemke, 2000).

This perspective has important implications for understanding teacher learning. Teachers derive socially-constructed and culturally-mediated principles of teaching from practice and reinstantiate these principles in future practice. Teachers first reify their past experiences as principles of practice that take the form of knowledge, beliefs, identities, and general orientations. These constructs do not exist as extant entities – rather, they are constructed and negotiated through activity and mobilized as tools in activity in light of norms and conditions of the contexts in which activities occur. Over time, their practice becomes more routinized. In teaching, it is routinized action that characterizes teachers' knowledge, demarcating the transitions from novice to expert teachers along the teacher professional continuum. This routinization is evidenced in a developed alignment between teachers' personal characteristics and features of their contexts, both material and symbolic (Barab & Roth, 2006; Brown, 2009). This developed alignment represents teaching expertise.

Questions surrounding teachers' knowledge, expertise, and learning have held a prominent role in education scholarship for many years (e.g., Putnam & Borko, 2000; Shulman, 1986). While such research has made invaluable contributions our understanding of teachers' knowledge, beliefs, identities, and orientations, a great deal remains uncertain about the nature of teachers' expertise and learning. Like many practitioners, teachers often struggle to articulate and represent the principles and rationales underlying their professional practice. There remains little to no consensus as to what exactly constitutes teacher expertise and how to differentiate between expert and novice teachers. We suggest that there exists a strong rationale for exploring the utility of novel theoretical and analytical perspectives in subsequent research on teachers' expertise and learning.

## Cultural-Historical Activity Theory (CHAT)

To better operationalize, capture, and describe expertise for teaching, cultural-historical activity theory (CHAT - Engeström, 1987) has been forwarded as a potentially useful tool for educational researchers (Grossman, Smagorinsky, & Valencia, 1999; Roth & Lee, 2007). Cultural-historical activity theory is a psychological, activity-based perspective on human activity and development. Consistent with situated perspectives on knowing and learning, CHAT affords a perspective on consciousness as an emergent property of interactions between groups of people in certain cultural contexts rather than an entity 'in the head'. As Engeström (1987) notes, it is a "concept of activity based on material production, mediated by technical and psychological tools as well as by other human beings" (pg. 25). In this way, CHAT is fundamentally concerned with socially- and culturally-mediated, as well as object-oriented, activity and the evolution of established practices over time.

The fundamental unit of analysis in activity theory is human social activity itself, driven by a goal-orientation relevant to a particular need or motive as defined by members of the community. Activity undertaken by an individual (subject) whose particular motive or need impels action oriented toward a particular problem or purpose (object). Consistent with the foundations of cognitive science laid by Vygotsky, such activity is also mediated by tools and artifacts (instruments) and by other human beings (community) within the activity system. The nature of activity as it develops is also structured and shaped by norms of the community (rules) and specialization or social stratification (division of labor). These complex relationships are embodied in the CHAT activity triangle, a generalized model for analyzing social activity, which is shown in Figure 1.

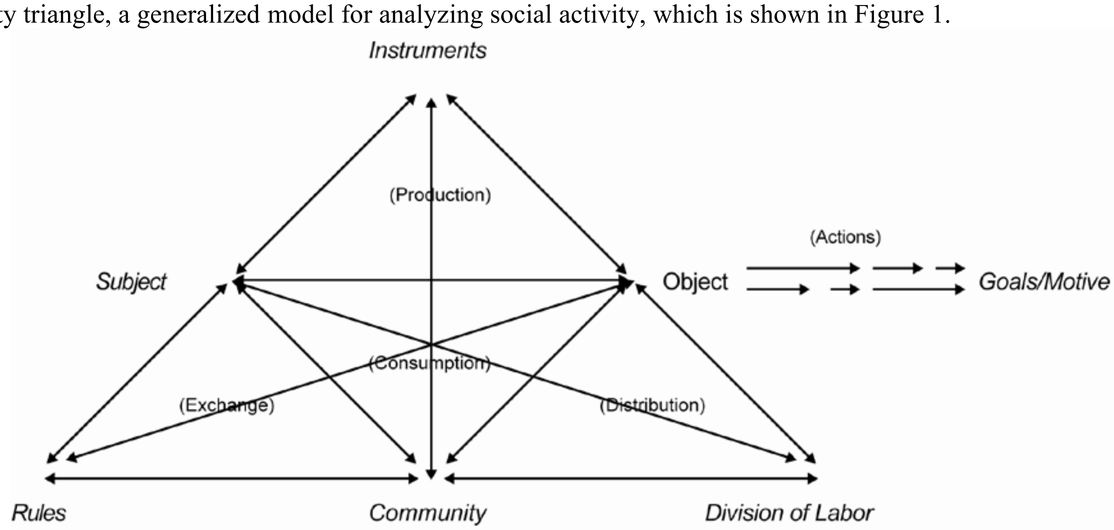



Figure 1. Cultural-historical Activity Theory Model of Human Activity (Engeström, 1987)

Activities are driven by their objects and motives. Activities are, however, also composed of composite actions which, in turn, are constituted by conditioned operations. Operations are best characterized as the kinesthetic foundation upon which sets of actions and, ultimately, activities, are based. These operations are the product of direct response to environmental stimuli, typically unconscious, routinized behaviors. In contrast, actions are consciously-driven by goals. Both actions and the goals toward which they are oriented are given meaning by the activities in which they are situated while, at the same time, constitute the achievement of the collective motive of activity.

The mediating influence of tools is a key assumption of activity theory. Vygotsky (1978) noted the importance of tools in transforming human activity from a direct to an indirect, or mediated, relationship to the environment. This is shown in the 'production' triangle in Figure 1, which illustrates the iconic triadic relationship between the individual, the environment, and mediating influence of available tools. Production processes, however, do not occur in the abstract. Rather, they are embedded in particular social and cultural contexts that serve to mediate and shape the relationships between individuals, their tools, and the particular focus of their efforts. Cultural and social influences on activity include three elements: rules, community, and divisions of labor.

The CHAT model in Figure 1 provides a complete account for the structure of a given activity system, or the central activity. However, third-generation activity theory assumes that the constitutive elements of a given central activity (i.e., nodes of the triangle) are themselves the products of related, interconnected activities. As such, a particular activity system is itself embedded in networks of activity systems. While these activities may appear stable, there exist ever-present tensions within and between nodes of activity systems and neighboring activity systems. These contradictions arise as "the clash between individual actions and the total activity system" (Engeström, 1987, pg. 30) and are the motor for and harbinger of change in activity. Contradictions are important because they lie at the heart of learning in practice. These contradictions, which develop within and between bounded activity systems or practices, like at the heart of learning in practice in that they must be negotiated and addressed to engage in culturally-more advanced forms of an existing practice. The goal of such learning is expansive development of new, more culturally-advanced and articulated forms of activity.

CHAT can serve as a useful tool to study teachers, teaching, and teacher learning (Grossman, Smagorinsky, & Valencia, 1999). CHAT centralizes activity itself as the fundamental unit of analysis, emphasizing goal- and object-oriented material production, the cultural mediation of these processes, and how a particular activity is nested within broader networks of systems. As such, it highlights the importance of learning in context (Putnam & Borko, 2000). Teachers learn from classroom practice but these are not the only settings in which they go about their work Especially in respect to preservice teachers, who traverse multiple activity settings on their way to becoming full-time, practicing teachers, accounting for these unique settings in which learning occurs is essential. Such a perspective also prioritizes the importance of tools, whether curriculum materials or others, that teachers use to structure and guide classroom practice. In this way, the CHAT model provides a mechanism through which to attend to both individuals and the worlds in which they learn and develop.

However, CHAT-based research on science teachers, teaching, and teacher learning is limited. In this paper, we present methods and findings from two parallel CHAT-based studies on science teachers (elementary and secondary) in professional learning contexts (formal teacher education and professional development). The specific purpose of this paper is twofold. First, we describe two studies that use CHAT to investigate science teacher learning through iterative cycles of instructional planning, lesson enactment, and reflection on practice. Second, we identify and discuss important theoretical contributions to perspectives on teacher learning and expertise that have emerged from these studies. Research methods, findings, and research-related issues discussed in this paper will not only help other researchers employ CHAT in research on science teachers, teaching, and teacher learning, but also inform theoretical discussions of the nature of teachers' expertise and learning.

## Methods and Main Findings

### Study A –Preservice Elementary Teachers' Development of Pedagogical Design Capacity for Inquiry-Based Science Teaching and Learning

Science curriculum materials remain one of the most widespread and important tools for teachers. However, rather than enacting them 'as-is', teachers actively mobilize, evaluate, critique, and adapt curriculum materials (Remillard, 2005). The curriculum design process is a function of teachers' personal characteristics (knowledge, beliefs, and identity, etc.), features of the curriculum materials, and features of their professional contexts (e.g., Roehrig, Kruse, & Kern, 2007). Together, these three factors, as well as the goals toward which classroom activity is oriented, constitute the pedagogical design capacity a classroom activity system affords a particular teacher (Brown, 2009). In order to leverage and maximize the capacity for pedagogical design afforded them, teachers need to learn to use curriculum materials effectively (Remillard, 2005). This is particularly important for preservice teachers who tend to rely heavily on curriculum materials as beginning teachers and who struggle to translate their reified understandings of inquiry-oriented teaching and learning into classroom practice. This study extends recent research on the use of curriculum materials in science teacher education (e.g., Forbes & Davis, 2008; Schwarz et al., 2008) by focusing on how preservice elementary teachers translate their espoused inquiry frameworks into planned and enacted science lessons, as well as how the curriculum design process is socially- and culturally-mediated across contexts.

This mixed-methods study involves 46 preservice elementary teachers in an elementary science teaching methods course and in-depth case studies of four elementary preservice teachers during the final year of their teacher education program. The focal point for data collection and analyses were science lessons the preservice teachers planned and enacted in elementary classrooms during the methods and student teaching semesters. Drawing on activity-theoretical frameworks, two interacting activity systems are articulated that are relevant to this study. Curriculum planning (central activity) and curriculum enactment (object-activity) are foregrounded as constituting curriculum design for inquiry, consistent with existing descriptions of teachers' practice (e.g., Remillard, 2005). This model provides a number of specific affordances. First, it highlights the use of both symbolic and material tools (preservice teachers' espoused inquiry frameworks and science curriculum materials), what Engeström (2007) refers to as 'tool constellations', with which they engage in curriculum design. Second, the model affords the ability to map the construction of science lesson plans as boundary objects between curriculum planning and enactment domains. Third, by identifying contradictions that the preservice teachers articulate within and across settings, it is possible to link their curriculum design decisions to underlying contradiction-specific rationales. Ultimately, this model, in addition to other representational tools described in the next section, affords the ability to trace the emergence and resolution of contradictions in curriculum design activity over time.

Data were collected throughout the academic year and included interviews, lesson plans, instructional artifacts, reflective journals, and observations of enacted science lessons. To code this data, a series of coding keys were developed that are explicitly aligned with the CHAT-based model used in this study. These coding keys afforded the ability to identify the preservice teachers' curriculum design decisions, assess the inquiry-orientations of their planned and enacted science lessons, and identify contradictions that emerged in their curriculum design activity. Using coded data from the total population of preservice teachers in the methods semester (n=46), quantitative analyses were performed to describe patterns in their curriculum design decision-making and to assess how inquiry-based their planned science lessons were. A regression model was also constructed to provide causal explanations for why their lessons were or were not highly inquiry-based. Qualitative methods were used to analyze data from the four preservice teachers studied over the year. These analyses were characterized by an iterative process of coding, reduction, displaying, and verification of data (Miles & Huberman, 1994) that were directed towards the development of case studies. The primary goal of these analyses was to identify contradictions within and across curriculum design contexts that explained the preservice teachers' curriculum design decisions. Multiple coding matrices were used to display these relationships, one that was lesson-specific and one that was contradiction-specific. The latter was necessary to trace the resolution of particular, pervasive contradictions over time and characterize preservice teachers' evolving inquiry frameworks and curriculum design practices.

Results from Study A illustrate how extant, reified models of classroom inquiry served as crucial instruments in the preservice teachers' curriculum design for inquiry. They developed their capacities to transform their professed models of inquiry into classroom practice by delimiting particular aspects of inquiry as the object of their curriculum planning activity. However, their curriculum design efforts led to emergent contradictions within their placement classrooms. The preservice teachers worked to resolve these contradictions in various ways, including constructing specialized, interim instructional plans; restructuring and reordering time; and reprioritizing particular facets of inquiry in ways that reflected these constraints. While the curriculum materials they used largely defined the material and conceptual space in which their curriculum design efforts occurred, it was the influence of their placement classroom contexts that ultimately determined how they were able to translate the abstract into the concrete. The preservice teachers ultimately represented the challenges they experienced enacting their science lessons as a fundamental contradiction between two object-motives of curriculum enactment focused on student learning. All were still struggling to reconcile this contradiction at the end of the study.

## Study B – How Pedagogical Content Knowledge Develops: The Impact of Reflection and Community

While many researchers have advanced the notion of pedagogical content knowledge (e.g., Shulman, 1986; Loughran et al., 2001), there remains a gap in our understanding of how this knowledge develops over the course of a teacher's career. There are many factors that likely influence the development of PCK, including the teacher's content knowledge within a subject domain, the students' prior knowledge, the pedagogical approaches employed, interactions between students and teacher, student assessments, and how the teacher reflects on these experiences before, during, and after the instruction (Morine-Dershimer & Kent, 1999; Magnusson, Krajcik, & Borko, 1999). Many scholars comment on the complex and cognitive nature of teacher planning and the influence of the teaching of any subject area on teacher knowledge (Magnusson, Krajcik, & Borko, 1999; Leinhardt & Greeno, 1986).

Teachers' knowledge growth can occur through professional development, which can "help teachers develop cohesive understanding about inquiry instruction by building on their existing ideas about student learning, technology and the role of the instructor" (Slotta, 2004, pg. 203). However, most teacher professional development is decontextualized. This study examines the role of reflection and peer-exchange in helping science teachers develop PCK though their design, enactment, and revision of a technology-enhanced, project-based lesson. An intervention is introduced where teachers reflect on their planning, enactment, and revisions, as well as with their peers in a community. CHAT is employed to analyze the connections among the activities in which teachers are engaged, the context in which these interactions are occurring, and the teachers' development of PCK.

This three-year longitudinal study uses an iterative design-based methodology to investigate the development of pedagogical content knowledge of nine science teachers (*n*=9) in relation to their instructional practices (e.g., lesson design, preparations, classroom interactions, assessment and feedback), and student understanding. The focus is on two specific interventions that serve to promote professional development: reflection, and peer-exchange. Teachers co-design and then enact a project-based, technology-enriched science lesson. Four main phases of teacher activities were captured by this study: (1) Prior teacher knowledge and experience; (2) Lesson design; (3) Classroom enactment; (4) Revision of lesson design.

Data sources include teacher surveys, interview questions, lesson plans, reflections (captured in a wiki), videotaped classroom enactments, field notes, student artifacts and responses, and peer exchanges (on wiki, and in group meetings). Following Koehler and Mishura (2005), all wiki documentation, interviews and field notes were coded for different categories of teacher knowledge, including pedagogical content knowledge. For each coded knowledge element, a qualitative score of 0-3 was assigned. For PCK, 0 represented the absence of that knowledge, while 3 represented a very clear understanding of students' prior knowledge and appropriate teaching strategies. These coded knowledge elements when combined with other enactment coding measures were used as data for subsequent CHAT analysis.

The activity and action patterns from the teachers' classroom enactments were identified from video documentation and field notes, and these were coded for types of action: Small group interactions (SGI); Large group interactions (LGI); Logistic actions (Log); and Isolated actions (Iso). Both SGI and LGI were subdivided into management (M) and pedagogical actions (Ped). Each type of action became its own segment. A qualitative score of 1-3 was assigned for the content of interaction. A score of 3 represented strong engagement and strongly addressing student learning needs. A score of 1 represented poor engagement or poor interaction. Students' activities were coded as either SGI or LGI and given a qualitative score of 1-3 based on the engagement, either with student peers or with teacher. A score of 3 represented high quality of engagement such as being on task and asking relevant questions. A score of 1 represents a weaker level of engagement. The SGI and LGI qualitative score was multiplied by the time frame of that segment and this value was then graphically represented. These codes were then cross-referenced to the nodes on an activity triangle for teacher enactment of project-based lessons, providing an iterative activity-based analysis. A within case and cross-case analysis of actions and operations (Miles & Huberman, 1994) provided evidence of challenges, contradictions and tensions that resulted in emerging knowledge or new learning opportunities for teachers.

Results in this study indicate that tools such as scaffolded wiki-reflections and lesson designs offered a mechanism for participants to make their actions visible to themselves and to their peers. The CHAT-based framework used in Study B helped expose tensions between subject-tool-object and community-tool-object relationships, which in turn affected the next iteration of the lesson design and teacher knowledge development, while also identifying the tensions within activity systems and the emergence of new rules. For example, during the lesson revision phase of the teachers' activities, participants were able to see specific flaws in the lesson design. All teachers had a pattern of lesson revision, which identified the student wiki-templates in the project-based lesson as underutilized. Through the reflective practices, teachers (subject) acted differently in their lesson design incorporating more scaffolding (rules) within the re-design of the lesson and, subsequently, enacted the lesson differently. In addition, teachers participating in the peer-exchange activities began to emphasize the structured use of wiki-templates as part of their lesson design. Tensions in the community-tool-object relationship caused dynamic shifts in the activity systems from one iteration to the next. Thus, the generative knowledge construction was not limited to self-reflections but also occurred through community exchanges and community reflections.

## Implications for Perspectives on Science Teachers' Expertise and Learning

There have been recent calls for increased use of CHAT in education research (Roth & Lee, 2007), including research on teachers' practice, expertise, and learning (Grossman, Smagorinsky, & Valencia, 1999). However, there remains little CHAT-based research focused on teachers and teachers' learning, particularly in science. The two parallel studies presented in previous sections are unique in that they draw explicitly upon CHAT as both theoretical and analytical frameworks to investigate science teacher learning. Findings from these studies reinforce the context-dependent nature of teachers' learning and expertise assumed in sociocultural theory (Putman & Borko, 2000). However, they also advance the field's theoretical understanding of how and why teachers' learning is situated by illustrating underlying contradictions that explain change over time in particular cultural contexts. In this section, we highlight important theoretical and practical implications for CHAT-based perspectives on teachers' expertise and learning that have emerged from these two studies.

A crucial element of CHAT is the break it makes with cognitivist psychology. Rather than characterizing learning as an internalization of mental schema or concepts, it assumes instead that learning is evidenced in changes in activity, emerging from the synergistic interaction between the symbolic (cognitive and cultural) and physical environment. CHAT foregrounds the translation of the abstract to the concrete, or realization of the symbolic in action through activity. However, the origins of expansive learning lie in the initial recognition of emergent contradictions and questions about the object and motive of a given activity. So, in effect, the beginning of the process of moving from the abstract to the concrete is spurred by the objectification of activity, or the articulation of ideas, principles, and truths from experience. Because activity also involves deriving the abstract from the concrete, they are therefore temporal, and the symbolic is already a product of the history of the activity in which it is mobilized and used. Ultimately, then, networked practices, as illustrated here in both studies with curriculum planning and curriculum enactment, rely not just on moving from the abstract to the concrete, but also from the concrete to the abstract through a cyclical process of internalization and externalization (Engeström, 2008), participation and reification (Wenger, 1998), and circulating reference between form and matter (Latour, 1999). What this suggests is that while practice-specific representations take many forms, they are both used *in* activity and developed *through* activity.

As a process of internalization and externalization (Engeström, 2008), activity systems appear twice to the subject, in this case the teacher; once in material form and once in symbolic form. Semiosis is an emergent property of both. For teachers, this highlights the dialectic by which they have, through activity, arrived at a particular point in their own learning, and has clear implications for perspectives on teacher learning. Most clearly, it highlights the need for teachers' to be engaged in teaching practice and for that practice to be foregrounded in programs designed to promote teacher learning. This is crucial not only for teachers to identify the contradictions and tensions that pervade classroom activity systems, but also for them to attempt to resolve those contradictions through activity. In Study A, for example, the preservice teachers articulated a fundamental contradiction in curriculum enactment for inquiry between two objects and motives of curriculum enactment. This tertiary contradiction emerged between positioning students as objects of activity who appropriate predetermined learning goals and as community members and co-collaborators involved in the co-construction of both the object and explanations of phenomena. This contradiction only emerged after engaging in multiple iterations of curriculum design for inquiry and was thus derived and objectified from activity. In Study B, teacher tools (such as online reflection notes, and lesson designs), highlighted teacher actions in classroom activities, and identified critical moments (tensions) where teachers recognized that a change in patterns of practice were required. The object in this study, PCK development, is connected to these tensions and contradictions. Through iterative activity-based analysis, these actions and changes in actions allowed the object in this study, PCK development, to be traced. These findings reinforce the importance of teachers' engagement in rigorous processes of self-reflection and analysis of practice.

From CHAT-based perspective, learning is defined as the evolution of an existing activity or practice to more culturally-advanced form of that activity. However, in addition to emphasizing teachers' learning as changes in practice, CHAT foregrounds the fundamental motives that drive activity and the physical and symbolic objects through which they are pursued. All activity, including science teaching and learning in the classroom, is situated. However, when cultural layers are removed from activities, they are left composed of the essentials of collaborative work – groups of individuals working together on a particular task that possesses shared meaning (Engeström, 1987). As a result, it is impossible to understanding teachers' learning without accounting for object-motives of their professional practice. In Study A, the preservice teachers' curriculum design for inquiry was punctuated by episodes of lesson-specificity determined by the curriculum materials they employed. While the episodic nature of their curriculum planning and enactment cycles resulted in lesson-specific tools and physical objects, their symbolic objects and motives largely remained constant – to promote students' construction of explanations from evidence through collaborative classroom inquiry. In Study B, the activity system occurs embedded in authentic teacher practice. The teachers design, enact and revise an innovative science lesson as part of their normal activity. However, active teacher learning requires collaboration with peers, as teacher knowledge development is demarked as a social process. The development of community and the role of reflection within the community are powerful elements for "expansive knowledge" (Engeström, 2007). Emerging teacher knowledge (object in Study B) is amplified through community exchange and community activity. As evidenced in both studies, to understand teachers' learning, it is critical to identify the object-motives that orient the activities through which their learning is manifested.

CHAT also emphasizes the material basis of activity, meaning that the products of activity are not just abstract concepts or ideas, but physical, tangible objects. Here, in both studies, the teachers' learning was made manifest in the curriculum materials they produced through iterative cycles of curriculum design. As boundary objects (Wenger, 1998) between curriculum planning and curriculum enactment, the teachers' lesson plans and other curricular artifacts served a dual purpose as a working problem space for curriculum planning and an instrument to support curriculum enactment. As such, they were representations of their learning, as others have argued before (Ball & Lampert, 1999; Loughran et al., 2001; Shulman, 1986). However, they also served as vehicles and conduits for the transport and application of expertise in novel contexts. For the preservice teachers in Study A, they were not only crucial instruments for translating inquiry frameworks into classroom practice, but also the contextualization and specification of those inquiry frameworks over time. In Study B, teacher lesson designs and student artifacts, manifested as evidence for teachers and for the peer-community. The teacher could comment on their learning in practice with these artifacts, and were able to articulate their meaning and their understanding of project-based learning. Given the challenges of promoting transfer, or learning across contexts, the focus on the physical and material basis of transfer inherent to CHAT is an important element of perspectives on teacher learning.

Finally, rather than emphasizing institutional stability, CHAT assumes that activity systems are inherently contradictory, full of tensions, and always subject to change. If such dynamicism exists as an integral feature of activity systems, including science classrooms, and it demands changes in practice to respond to emergent contradictions, then teacher learning from a CHAT-based perspective is a foregone conclusion. But this does, perhaps, speak more specifically to perspectives on teachers' expertise. Teachers' expertise must not only be marked by their ability to operate within particular activity systems, but also to effectively respond to changes to those activity systems and emergent contradictions. For the preservice teachers in Study A, the introduction of inquiry as an extant framework necessitated changes in their teaching practice, though likely made less contradictory by their lack of experience and participation in a mature, preexisting form of teaching practice. What came to define their expertise for inquiry-based science, however, was the ability to respond to unpredictable schedules (rules) and curricular resources (instruments) to engage students in inquiry practices. In Study B, teachers activity systems are dynamic, fluid and in constant flux as they are dependant on the enactment of the lesson. This enactment addressed school schedules, student learning and content-discipline. The revision in lesson design (artifacts) and rules that were set up for the next iteration of enactment occurred as a response to the tensions within the original activity system. Often these shifts or changes were first recognized within the classroom (or on the fly), prompting a quick adjust during the enactment which became more formalized in the revision of lesson design for the next iteration.

## Conclusion

The two studies presented here contribute to a limited but growing body of CHAT-based education research. Specifically, the two studies described in this paper illustrate how CHAT can be employed in research on science teachers' expertise, practice, and learning. As shown in the discussion of these two studies, such research can and, we believe, will continue to yield novel insights into science teachers and teaching, both empirical and theoretical, which possess the capacity to inform a multitude of design efforts focused on supporting science teachers' practice and development of expertise. However, the fusion of the CHAT framework with the material and conceptual spaces in which teacher learning and practice takes place is not without its struggles. As an emergent research domain, it presents many questions to yet be addressed, particularly methodological ones. We believe that these are important issues that should be further explored for the promise of CHAT to be more fully realized in educational research.

## References

Ball, D.L. & Lampert, M. (1999). Multiples of evidence, time, and perspective: Revisiting the study of teaching and learning. In E.C. Lagemann & L.S. Shulman (Eds.), *Issues in education research: Problems and possibilities* (pp. 371-398). San Francisco: Jossey-Bass.

Barab, S.A. & Roth, W-M. (2006). Curriculum-based ecosystems: Supporting knowing from an ecological perspective. *Educational Researcher, 35*(5), 3-13.

Brown, M. (2009). Toward a theory of curriculum design and use: Understanding the teacher-tool relationship. In J. Remillard, B. Herbel-Eisenman, and G. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17-37). New York: Routledge.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research.* Helsinki: Orienta-Konsultit.
Engeström, Y. (2007). Enriching the theory of expansive learning: Lessons from journeys toward coconfiguration. *Mind, Culture, and Activity, 14*(1-2), 23-39.
Engeström, Y. (2008). From design experiments to formative interventions. In *Proceedings of the 8th International Conference of the Learning Sciences* (pg. 3-24). Utrecht, the Netherlands: ISLS.
Forbes, C.T. & Davis, E. A. (2008). The development of preservice elementary teachers' curricular role identity for science teaching. *Science Education, 92*(5), 909-940.
Grossman, P., Smagorinsky, P., & Valencia, S. (1999). Appropriating tools for teaching English: A theoretical framework for research on learning to teach. *American Journal of Education, 108*(1), 1-29.
Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research*, 32(2), 131–152.
Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies.* Cambridge, MA: Harvard University Press.
Lienhardt, G. & Greeno, J. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*(2), 75-95.
Lemke, J.L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture, and Activity, 7*(4), 273-290.
Loughran, J., Milroy, P., Berry, A., Gunstone, R., Mulhall, P. (2001). Documenting science teachers' pedagogical content knowledge through PaP-eRs. *Research in Science Education,* 31, 289-307.
Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95-132). The Netherlands: Kluwer Academic Publishers.
Miles, M.B. & Huberman, A.M. (1994). *Qualitative data analysis.* Newbury Park, CA: Sage.
Morine-Dershimer, G. & Kent, T. (1999). The complex nature and sources of teachers' pedagogical knowledge. In J. Gess-Newsome & N.G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 215-234). The Netherlands: Kluwer Academic Publishing.
Putnam, R.T. & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4-15.
Remillard, J.T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*(2), 211-246.
Roehrig, G.H., Kruse, R.A., & Kern, A. (2007). Teacher and school characteristics and their influence on curriculum implementation. *Journal of Research in Science Teaching, 44*(7), 883-907.
Roth, W-M & Lee, Y-J. (2007). "Vygotsky's neglected legacy": Cultural-historical activity theory. *Review of Educational Research, 77*(2), 186-232.
Schwarz, C., Gunckel, K., Smith, E., Covitt, B., Enfield, M., Bae, M., & Tsurusaki, B. (2008). Helping elementary pre-service teachers learn to use science curriculum materials for effective science teaching. *Science Education, 92*(2), 345-377.
Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.
Slotta, J. D. (2004). The Web-based Inquiry Science Environment (WISE): Scaffolding knowledge integration in the science classroom. In M.C. Linn, P. Bell, and E.A. Davis (Eds.), *Internet environments for science education* (pp. 203-232). Mahwah, NJ: Lawrence Erlbaum.
Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes.* London, England: Cambridge University Press.
Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* New York: Cambridge University Press.

## Acknowledgements

This research is supported by the National Science Foundation (grant numbers REC-0092610, CLT-0227557, and DRL-0733299), the University of Michigan Rackham School of Graduate Studies, and the Ontario Institute for Studies in Education at the University of Toronto. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors.

# Using conceptual blending to describe emergent meaning in wave propagation

Michael C. Wittmann, University of Maine, 5709 Bennett Hall University of Maine, Orono ME 04469-5709, wittmann@umit.maine.edu

**Abstract:** Students in interviews on a wave physics topic give answers through embodied actions which connect their understanding of the physics to other common experiences. When answering a question about wavepulses propagating along a long taut spring, students' gestures help them recruit information about balls thrown the air. I analyze gestural, perceptual, and verbal information gathered using videotaped interviews and classroom interactions. I use conceptual blending to describe how different elements combine to create new, emergent meaning for the students and compare this to a knowledge-in-pieces approach.

## Introduction

In several strands of research into student understanding of specific science content, one research assumption is that ideas (be they large scale (Wandersee, 1994; Carey, 1985) or small scale (diSessa, 1993; Hammer, 2000)) already exist and are triggered in a given context, ready to be used, fully formed. The ideas are given various names within physics education research, either misconceptions such as "impulse theory" (McCloskey, 1983a,b), phenomenological primitives (p-prims) such as "force as mover" (diSessa, 1993), or reasoning resources such as "actuating agency" (Hammer, 1996). Across all these different traditions, the research involves modeling which pre-existing ideas are activated and how they are used with each other in a context.

Ideas are not only activated. New ideas must come into existence, else learning would not occur and our toolbox of useful ideas would never increase. This prompts the need for explanations of the development of ideas, be they thought of as conceptions, p-prims, or resources. One could assume a long-term development of an idea through many processes. A rich literature on the development of misconceptions (see Wandersee, 1994, for a review) suggests that misconceptions are formed through inappropriate teaching, misinterpretation of observations of the world around us, or describing situations using inappropriate models. Primitives, on the other hand, might develop very early in life (Mandler, 1992), suggesting that some kinds of p-prims are nearly hard-wired into our thinking. We might reify the several processes we use to solve a problem into a single new cognitive object (Dreyfus 2002; Sfard, 1991; Tsamir, 2004). We can also think of connecting many different resources into a consistently activated network appropriate for a given setting (Hammer et al., 2004) or we can discuss ways in which networks of resources rearrange so as to describe different types of conceptual change (Wittmann, 2006). The plasticity of such networks can be defined, as well (Sayre, 2008). In summary, one can consider many pathways to the development of robust ideas which are later able to be activated.

Not all ideas need be developed in a robust fashion and for long times. Indeed, it is clear that new ideas must arise in new and novel situations. Such ideas might develop on the time scale of seconds and last for only as long as a problem needs solving. Thus, it is appropriate to ask if we can observe such events occurring: do we see students using an idea on very short time scales, and, if so, what observations allow us to conclude so?

In this paper, I suggest one approach to modeling emergent meaning in students who are solving physics problems. I look at common students' responses to a typical physics question about wave motion on a long, taut spring. Students were asked to describe how one creates a single wavepulse on the spring, and then predict how one might change the speed of the wavepulse. I show how students commonly use similar language and gestures as they give their answers. In summary, their answers are similar to descriptions given to describe how a ball is thrown through the air. To model the details of their answer, I use a certain level of formalism that is borrowed from the field of language development and linguistics.

I use the mechanisms of mental space integration, also called conceptual blending (Fauconnier and Turner, 2002), in particular the ideas of composition, completion, and elaboration. In this analysis, I consider two different mental spaces, one being the observed event of a wavepulse being created by a wrist flick and traveling along a long, taut spring, the other being the imagined space of a ball thrown in the air. These two spaces are connected by similar elements, forming a blend composed of elements of both spaces. New information is recruited from the different mental spaces, but only some information from each space is used. The idea represented in the blend is then elaborated upon, leading to emergent meaning. Specifically, students predict that flicking one's wrist harder leads to a faster wavepulse (as if throwing a ball harder). This response seems obvious within a simple metaphor (a wavepulse is like a ball) but gives us insight into ways in which student responses emerge rather than are activated. In particular, one well-documented idea, the Ohm's p-prim (diSessa, 1983, 1993) can be thought of as emerging from the blend rather than being a pre-existing idea that is activated in a new setting. This result is consistent with observations of students engaged in a time-consuming construction of new meaning as they explain their thinking.

## Student responses to a wave propagation problem

The data analyzed in this paper come from interviews where students make predictions about a basic physics problem that is commonly taught and studied in a high school or college survey physics course. Data presented in this section are taken from the author's 1998 Ph.D. dissertation, and have previously been analyzed in several ways (Wittmann 1999, 2002). Data on the question discussed here were gathered from hundreds of students in written questions (free response, multiple choice, and multiple-choice multiple-response questions on ungraded quizzes and specially designed, ungraded pre- and post-instruction surveys, as well as graded tests), tens of interviews, and hours of informal classroom observations. Rather than analyze all students, I am interested in a class of common responses for which I will give a few brief examples. The examples below illustrate the most typical response given by students. Data come from transcripts from interviews carried out with students who were made aware of the Think Aloud protocol (Someron, 1994). The role of the interviewer was to elicit responses about hypothetical situations and explore the reasoning students gave for their responses. Though the original intent of the interviews was not to explore the development of ideas, the data lend themselves to a re-analysis in terms of emergent meaning arising in new and novel situations.

The physics problem concerns wavepulse propagation along a long, taut spring. In a typical form of the interview question, students were given an image of an already propagating wavepulse (see Figure 1). They were asked to describe how such a wavepulse could be created. Then, they were asked how they might change the amount of time it took for the wavepulse to reach the wall; some students were asked to increase the time, some to decrease it. A correct response to such questions would have been to consider the tension in the spring (changing how tightly the spring is pulled) or changing the mass density of the spring (by replacing the spring with a different one, basically). The hand motion used to create the wave does not affect the speed of propagation. (A complete analysis of this situation is found in typical introductory physics textbooks and was part of instruction that all students had completed by the time of these interviews.)

Figure 1. Propagating Wavepulse. A hand moves up and down and creates a pulse which moves toward a distant wall.

Three examples illustrate common features of many student responses. "Ford" (all names are aliases) had already demonstrated how to create the wavepulse, making a brief up-and-down motion with his hand to show how the shape came to be. The following dialogue occurred shortly after, with information about his drawings and gestures included parenthetically:

IVW: Let's say that you are standing over there, you create it, and it takes a certain amount of time for it to reach the wall. Let's say you want to shorten the amount of time that it takes. What would you have to do?

Ford: One -- I could probably -- there are two scenarios that I have to think about, and since you want me to say right now -- I'd send a quicker one (he draws a pulse smaller than the original in both width and amplitude and makes an up-and-down hand motion in less time and with lower amplitude than the original)

I: Wait, let me stop you right here, by quicker you mean, you did your hand motion like this (copies Ford's small-amplitude motion) --

F: Shorter, I wouldn't go (repeats original large-amplitude hand motion) I'd try to make a shorter hand motion (repeats small-amplitude hand motion), I wouldn't want to like this (repeats original large-amplitude hand motion). It would get there faster.

Notably, Ford's explanation of "I wouldn't go" involved a grammatical pointer to his hand motion, the gesture acting as the predicate to the sentence. This first explanation was immediately followed by a second scenario in which Ford described a larger and therefore faster pulse instead, again with associated hand motion. This dialogue was long and is left out because essential features of the explanation have already been presented – the differences in hand motion and language pointing to them.

"Adam," when asked a similar set of questions, had given the same kind of explanation for wavepulse creation, a flick of the wrist up and down. His explanation used slightly different physics terminology, but involved the same sense of a change to how the wavepulse is created:

IVW: Imagine that we can measure, just clock the time it takes the pulse to go all the way down to that wall. How could you shorten the amount of time that it takes for it to reach that wall, what could you do to make the time less?

Adam: If we could make the initial pulse fast, if you flick it, you flick it faster.

I: ...And why would flicking it faster make it go faster?

A: That would make it, I don't know, it's kind of hard to explain
I: Do the best you can
A: It would put more energy in or something.

Like Ford, Adam emphasized the connection between the hand motion and the speed of the wave. Another student, "David," gave a similar response to Adam and Ford's when he said "I think possibly, you see a slower pulse … if the force applied to the spring is reduced ... that is: the time through which the hand moves up and down [is reduced]." Again, the "force applied to the spring" describes how the hand moves up and down and affects the subsequent speed of the wave.

Many students' explanations were strongly connected to the hand motions they made to indicate the creation of the original wavepulse. The "faster flick" or the "greater effort" or the "smaller wave" were all shown through gesture, typically a harder hand jerk, a quicker flick of the wrist, a more vigorous and robust arm movement, and so on. Their motions and words were consistent with the idea that they were thinking as if the wavepulse were like a ball. Formalizing this intuitive sense of their responses is the purpose of the next section.

## Summary of Conceptual Blending

My goal is to account for the ways in which students used hand and arm gestures and a specific language to arrive at a certain kind of prediction about how to change the speed of the wave (involving changes in effort in creating a wave). It would be facile to claim that they activated the idea that more effort leads to more speed in the face of resistance of the string to being moved (an application of the Ohm's p-prim (diSessa, 1983, 1993) in the context of wavepulse propagation). But, the data suggest that one can think of the idea not being activated (as Ohm's p-prim) but as emerging in this context. To describe emergent ideas, I will use the formalism of mental space integration (commonly called conceptual blending), a framework developed by Fauconnier and Turner (2002) to account for meaning generation in language. Briefly, mental space integration assumes that there are different mental spaces (ideas associated with a given situation) that can combine in specific ways to create new meaning. Distinct mental spaces are combined due to some shared content or structure. The two spaces are brought together ("blended"), with selective projection taking some information from each input to compose a blend. New structure and new information is recruited (perhaps from long term memory) to complete the blend. One can let the blend "run", i.e., let the newly developed idea be elaborated upon. Emergent meaning arises as the results of this elaboration are connected by backwards projection from the blend back into at least one input space, perhaps both. We observe all these elements in student responses to the wave physics question.

I should be clear that I do not consider the analysis of student responses to be a linguistic problem, but instead am using representations and mechanisms from mental space integration to analyze reasoning in physics. I present applications and explanations in the wave physics context below; more detail can be found in books (Fauconnier and Turner, 2002) and in special issues of the journals of Cognitive Linguistics (as summarized in Coulson and Oakley, 2000) and the Journal of Pragmatics (as summarized in Coulson, 2005). The discussion that follows is particularly indebted to Bache (2005) and Hougaard (2005). I introduce the idea of blending with an example that summarizes many of the ideas used later in this paper to describe reasoning about wavepulses. It is adapted from Fauconnier and Turner (2002).

Consider a situation in a lab where you and a careless colleague are setting up an instrument. Your careless colleague is across the room, and about to attach one piece of equipment to another in a way that will cause trouble. You blurt out, "If I were you, I wouldn't do that!" Your colleague stops. What's going on here?

We can imagine two input spaces. You (and what you would do) exist in one space, your colleague in the other. In the blended space, the two of you are fused into a single, unique person ("if I were you" turns into "I am you"). This process of creating a new entity (of you and colleague as one person) is a "composition," in that it is composed of elements from both inputs. This new individual has an agency that is not yours (it is your colleague's) but makes a decision that is not your colleague's (it is, instead, yours). You have knowledge of the system that your colleague does not (which guides your statement "I wouldn't do that!"). This information "completes" the blend. The projection of information from each input space is selective, in that the judgment from your space is used for the decision in your colleague's space. In the blend, your knowledge is applied to you-and-colleague-as-one. You "elaborate" on the knowledge by implying that something should not be done. New meaning emerges – there is danger here. From here, there is back projection, from the blend back into your colleague's mental space: your actual colleague (thankfully) acts differently than originally intended.

To summarize important terms from this example, there are three major aspects that drive blending:

- Composition: creating relations that are not necessarily obvious, but allow the blend to occur
- Completion: recruitment from long term memory, for example in pattern completion or assigning properties to blends based on knowledge from one input
- Elaboration: "running the blend;" imaginative mental simulation that leads to emergent meaning that was not necessarily part of either input

Further terms of importance are selective projection (not all elements of each input space go into the blend, since often there are contradictory elements in the input spaces) and backward projection (information going back from the blended space to at least one input space). The full formalism of mental space integration is much larger than these five elements and includes heuristics for how distinct mental spaces are connected. That formalism is helpful in modeling student reasoning and provides additional details (such as the concept of the compression of cross-space connections), but is outside the scope of this paper. In the next section, I apply the formalism to the data presented above.

## Conceptual Blending of Wavepulses

The formalism of conceptual blending allows us to take student responses (including language and gesture) and make sense of their responses as emergent meaning, without activation of pre-existing ideas. I describe two individual inputs, show how they blend into a single space, and describe the emergent meaning that arises in the blend. I give two examples, the first based on the data that we have of student responses to the interview question, the second a hypothetical discussion which shows that a different set of inputs might lead students to create a blend in which the correct answer emerges. In the following section, I compare this analysis to past analyses (including my own) using another model of student reasoning.

### Seeing waves as balls

One can describe two mental spaces in use when predicting how to change the speed of a wavepulse translating down the spring. First is the simple perceptual observation of the wavepulse, a bump on a spring created by a vigorous hand motion and moving down the spring. Second is one of balls or other finite sized things that can be thrown through space. An understanding of object motion happens extremely early in life (Mandler, 1992) and ideas about thrown balls are readily available to most children (and university students). How these ideas get connected to the wavepulse is described by mental space integration. The result is a blended space with emergent meaning: wavepulses “as” balls, with certain properties that are not there in the observation itself but play out in predictions, when the blend is “run.”

A blend is typically described by a diagram such as shown in Figure 2. In the upper left, there is an input related to the observed wave on the spring. As shown by the students, there is a kinesthetic (proprioceptive) element of the wrist flick. One observes a pulse, as well as its motion (translation) down the spring. In the upper right, there is an input that contains typical properties of thrown balls. The representation takes into account issues of selected projection from each input into the blend. The Observed Wave input contains only a selective set of the observed phenomena, while the Ball input contains only a selective set of properties associated with balls. Only those elements of the Ball input which match to elements of the Observed Wave input are shown.

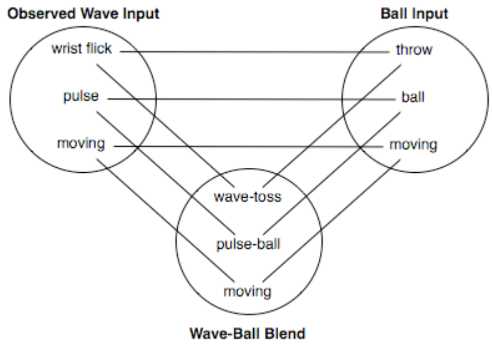



Figure 2. Wave-Ball blend. Two inputs (of the observed wavepulse and of thrown balls) are integrated to create a blend that leads to emergent meaning.

The Observed Wave and Ball input spaces are connected by like elements and integrated to create the Wave-Ball blend. There is selective projection, implied already because of the presentation of only selected pieces of information in each input. The wrist flick (perpendicular to the direction of motion) and a throw (in the direction of motion) are connected because they are both motions that initiate movement. The observation of a finite size pulse is connected to that of finite size objects, perhaps through a perceptual process of *binding*, a psychological term that describes how objects are seen as a whole. The act of binding is impossible for the (healthy) mind to stop. The motion of the pulse down the spring and a thrown ball through the air are connected because they are both propagation in a specific direction. Certain information is ignored: the perpendicular motion of pieces of the spring, for example, and information about the parabolic path of a thrown ball. It is probably unavoidable, based on the creation (by hand flick), shape (finite size), and directed movement of the wavepulse, to *not* think of the wavepulse as a finite-sized object like a ball.

In the blend, one can recruit information from the Ball input, namely that the speed with which you move your hand affects how fast a ball moves through the air. Furthermore, student predictions about how to change the speed of the wavepulse require that they "run" the blend. This leads to emergent meaning that was not present in the Observed Wave input. Notably, the hand motion perpendicular to the direction of wavepulse motion is taken as analogous to the throw of a ball in the direction of motion – obviously, they are not the same, but they are treated as such in the blend. So, in the blended space, the speed with which you move your hand affects the speed of the wavepulse along the spring; students like Ford, Adam, and David either describe a "quicker, faster hand motion" or that you "put more force in your hand" (with an accompanying larger/harder/faster wrist flick) when giving this answer. In the blend, the idea emerges that greater effort leads to greater speed. The Ohm's p-prim is not activated, it emerges.

## Seeing waves as events

One instructional goal of teaching students about waves has been suggested as helping students see the wave as a propagating event rather than an object translating down the spring (Hammer, 2000). For mental space integration to be useful, it must also describe what we believe expert physicists are doing when they think about waves. In Figure 3, we propose a blending diagram to describe how a physicist might model this situation.

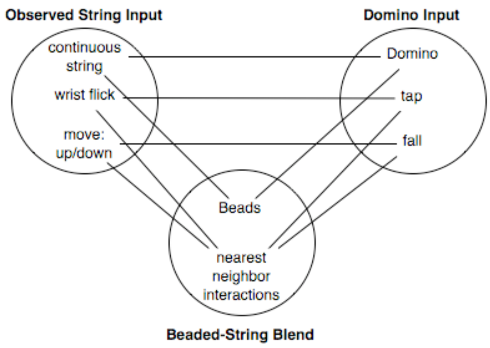



Figure 3. Wave-as-Event blend. In the blended space, the propagation of a wavepulse along a beaded string is an emergent phenomenon.

The blending diagram for this situation again contains two input spaces and a blend. The scale of observation is changed – rather than an Observed Wave (where one focuses on the propagating bump on a long, taut spring), one looks at the spring and has an Observed Spring Input. Three observations can be made. First there, is a long, continuous (taut) spring. Second, there is a flick of the wrist that is holding the end of the spring. Finally, one observes that the spring moves up and down (assuming, as always in this paper, a transverse wavepulse as shown in Figure 1) first at the end, and then consecutively at locations further down the spring. In the upper right of the figure, I suggest a second input: Falling Dominoes. It is important to note that any input containing effects spreading from discrete points to discrete points is sufficient. Examples include news spreading through gossip or a pile-up in a car crash during a traffic jam. The Dominoes example has the benefit that it is part of most people's everyday experience, is easily imagined, and is linear. In the Dominoes input space, there are individual dominoes, lined up in rows. One is tipped over, hitting the next. They fall down sequentially, each being knocked down by the one before it.

The blend is created through several connections. The spring and the Dominoes are both in a long line, so are visually similar. The motion of the wrist up and down at a single point (the end) of the spring is connected to the motion of the first Domino as it is tipped over. Just as each piece of spring goes up and down as it is pulled by the piece of spring next to it, the Dominoes keep falling over and hitting the next Domino in line. We have three elements: the long line, the repeated motions, and propagation of these events along the line.

The blend is composed of elements from each input. In the blend, the continuous spring is re-interpreted as being made up of discrete elements – the structure of the line of Dominoes is applied to the spring. Such a description was already implicitly used in the preceding paragraph, where "each piece of spring" acts on the next (this is not actually true in a continuous system). So, we have a Beaded-Spring in the blend. It moves according to the rules of the Observed Spring, where pieces of the spring go both up and down rather than simply falling over like Dominoes do. The blend, then, is a way of imagining the continuous spring as a set of discrete points that interact with each other through nearest neighbor interactions (like the discrete Dominoes) and move like the observed spring (rather than the discrete Dominoes).

The wavepulse is a global, emergent phenomenon of the blend. It arises because of elaboration, in which one "runs" the blend and plays out the consequences of the wrist flick and nearest neighbor interactions. The up motion of the wrist creates a signal down the spring; nearest neighbor interactions create the leading edge of the wave. The wrist changes direction; the peak of the pulse is created and propagates through the system as each bit of spring recapitulates the motion of the wrist. The signal ends when the wrist returns to its starting position; the wavepulse is completed and propagates through the system. The actual rules of propagation are not observed and only emerge in the blend. The speed of wave propagation depends on the nearest neighbor interactions due to tension and the mass of each individual “piece” of spring. We note that a typical textbook derivation of the wave equation describing wave propagation shows exactly this decomposition of a continuous spring into discrete segments and interactions among the nearest neighbors as the mechanism for propagation

What separates the Beaded-Spring blend from the Wave-Ball blend is the selective attention to the physical system being observed. It should come as no surprise that one reasons differently when attending to different details of a situation. Hougaard (2005) and Bache (2005) describe the choice of what to attend to as a process of disintegration, as one determines which parts of a system to use in the process of selective projection into the blend. In both cases, a motion acting on a system is connected to a different motion acting on a different system. One’s choice of system, balls or Dominoes, suggests different ideas of how to make sense of an observation. I do not pursue the analysis further in this paper, except to point out that choosing to observe either the “wave on a spring” or the “spring in motion” allows for different ways for dis-integrating the observation into blendable pieces.

## A knowledge-in-pieces analysis, instead

The analysis given above is not the only analysis possible with the simple interview data presented earlier. In this section, I review a previous analysis of wave propagation on a long, taut spring (Wittmann, 2002). Another possible analysis might be to use elements of the resources framework, a knowledge-in-pieces schema model of reasoning that builds off of diSessa's work (1983, 1988, 1993) on phenomenological primitives and is described more generally by Hammer (1996, 2000, 2004). Many of the ideas of the resources framework are consistent with Marvin Minsky's descriptions of frame systems (1975) and agents (1985). A more general review of the resources framework as used in this section is provided elsewhere (Wittmann, 2006).

### Resources as knowledge pieces

Resources are ideas that are useful and productive when solving some problem (Hammer, 2000). They are basic ideas one has that apply to a situation; they can be thought of as individual, nestable bits of knowledge.

As an example, “closer means stronger” (Hammer, 1996) implies that if you sit closer to a loudspeaker, the sound is louder; if you sit closer to your loved ones, you love them more; if you are closer to the sun, the heat is greater and it's summer. The first is typically true, the second of indeterminate veracity, and the last is false. The resource “closer means stronger” has no inherent rightness or wrongness. It's simply an idea, a knowledge bit, applicable (or not) in a given setting, and useful (or not) when engaged in problem solving. The term “resource” was originally meant to be very general, in the sense of computing, where a resource can be a printer (a tool outside the computer but useful for a program inside the printer), an API (in the software), or a re-usable object (in the code). One thinks of resources as activated or not. Typically, authors have been unclear about the origin of resources (whether they are built on the fly or exist as pre-compiled pieces ready to be activated), suggesting that both may be the case. In this section, for the sake of comparison with the previous discussion of blending diagrams and emergent ideas, we assume that they are pre-compiled and activated.

Few problems are solvable using a single resource; typically, many are necessary. Resources might coordinate with each other, as described by coordination classes (diSessa and Sherin, 1998). Furthermore, resources are useful at different scales - some very complicated concepts can act as resources in a problem solution (e.g., “force” as a primitive in the sense of “force as mover” (diSessa, 1993)) but can also be thought of as being made up of many individual resources (e.g., the coordination of resources and readout strategies about when to use them, described by the coordination class “force” (diSessa, 1998)).

### Resource graphs of wave physics reasoning

I rarely think of students (in interviews or classroom discussions) as thinking in terms of large-scale concepts (Wittmann, 2006). Instead, a simple model of students assumes that they activate only those parts of a concept that are relevant to the situation they are discussing. Similarly, I rarely expect that they use only one resource in a context, since a problem typically requires linking together more than one resource. I refer to the in-between level of description, neither primitive nor concept, as a mesoscopic description and use resource graphs (Wittmann, 2006) to represent this space of student thinking. A resource graph (see Figure 4) is a simple graphical representation of the resources activated by an individual in a particular setting.

Three resources, two previously described in the literature, can be used to account for student responses. First, and simplest of all (but important for later analysis) is Motion, specifically the translation of a bump along a long, taut string. This translation is caused by some startup effect, an Actuating Agency (Hammer, 1996), which diSessa has called Force as Mover (diSessa, 1993). Finally, increased effort in the face of resistance of the spring to being moved leads to greater speed, an example of the Ohm's p-prim (diSessa, 1983, 1993). These three resources are represented in Figure 4.

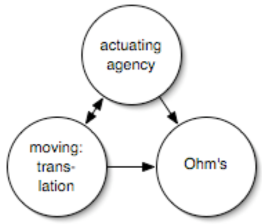



Figure 4. Resource graph to describe the typical response of thinking of wavepulses like balls.

Much as both the common student response and a hypothetical correct response could be modeled through a blending diagram, one can represent a hypothetical correct response using a resource graph, as well. As with the resource graph of waves-as-objects, the resource graph of waves-as-events contains only three elements (see Figure 5). There is an Actuating Agency that is, this time, carried out specifically on the spring. There is a displacement away from and back toward equilibrium for the end of the spring. Through Nearest Neighbor interactions, each element of the spring acts on its neighbor as it was acted upon. Finally, there is Motion, namely the propagation of the up-and-down motion of the hand.

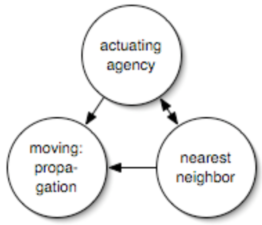



Figure 5. Resource graph for waves seen as events.

## Comparing Blending Diagrams to Resource Graphs

Mental space integration provides a more complete description of student thinking than does the resources framework because it makes fewer assumptions and has greater explanatory power.

There are fewer assumptions made about the nature of pre-existing ideas such as the Actuating Agency resource and how it is applied to different systems. In the blends, the connection between hand motion and the creation of the wavepulse or the motion of the spring is clear. In the resource description (as summarized in the graphs in Figures 4 and 5), one must make more assumptions about how the idea is applied – why was Actuating Agency connected to the wavepulse or to the spring? The use of the resource in its specific context is underdefined within the resources framework, while mental space integration shows directly how an idea is used in a setting.

Mental space integration also gives added explanatory power, indicating how the Ohm's p-prim emerges in a situation that is novel to students (as shown by their seeking many different explanations for their predictions). We need not assume that the Ohm's p-prim already exists and is activated in this setting. Instead, we can describe that the idea emerges, without assumptions of its pre-existence, and can be specific about what aspects of the blend led to its emergence. I do not doubt that there are cases where the Ohm's p-prim is activated as a pre-existing idea, but suggest that in this situation, it emerges. I generally believe that students have a self-checking mechanism ("does this idea make sense?"), and believe that they apply this and find some consistency between the Ohm's p-prim and their emergent idea. There is no evidence of such self-checking, though. It may be that the Ohm's p-prim acts as a "primary metaphor" to guide student thinking about a problem (Grady, 2005) and requires no self-checking mechanism. This suggests a way to combine the resources and blending analyses.

The idea of emergence within a specific context helps clarify some ambiguity in the resource graphs shown in Figures 4 and 5. Motion as a resource was related either to the translation of an object-like wavepulse or the propagation of a wavepulse; the blending description gives greater detail. In the Wave-Ball blend, the wavepulse moves like a ball, and a greater hand motion causes a greater wave speed. In the Beaded-String blend, the wavepulse emerges from elaboration of the blend. These are two very different descriptions.

Finally, modeling student responses with mental space integration provides a predictive power that is not possible in the resources framework. For example, if one throws a rock in a pond, water waves travel outward. An actuating agency caused a motion – but would one predict that a larger rock would cause a faster motion? Yes, the splash is larger; no, the speed is not. The resource graph in Figure 4 does not account for such detail, while the blending diagram in Figure 2 helps us recognize that the two situations might not be connected at all. Without a hand motion, the ball-space ideas simply won't be connected to the water-wave situation and it seems reasonable to predict that students would not think that a larger splash leads to a faster wave in the pond. Similarly, even though a hand is needed create a sinusoidal wave, its motion is not the same wrist-flick as when creating a wavepulse, and the idea of a throw might not exist to connect the one space to the other. Indeed, we find that few students talk about flicking the wrist harder when predicting how to increase the speed of sinusoidal waves.

Modeling student reasoning using mental space integration does have inherent problems. The description is more difficult, including issues of composition, completion, and elaboration (as well as details only alluded to in this paper, such as the compression of cross-space connections). In its favor, the assumptions about students' pre-existing knowledge are fewer, the explanatory power is greater, and the approach allows for a finer grain analysis of what is, and is not, affecting student reasoning in a given context.

## References


Bache, C. (2005). Constraining conceptual integration theory: Levels of blending and disintegration. *Journal of Pragmatics*, 37:1615–1635.

Carey, S. (1985). *Conceptual change in childhood*. The Massachusetts Institute of Technology Press, Cambridge, Massachusetts.

Coulson, S. and Oakley, T. (2000). Blending basics. *Cognitive Linguistics*, 11(3/4):175–196.

Coulson, S. and Oakley, T. (2005). Introduction. *Journal of Pragmatics*, 37:1507–1509.

diSessa, A. A. (1983). The ohm's p-prim. In Gentner, D. and Stevens, A. L., editors, *Mental Models*. Lawrence Erlbaum Associates, Hillsdale, NJ.

diSessa, A. A. (1988). Knowledge in pieces. In Forman, G., Pufall, P. B., and et al., editors, *Constructivism in the computer age.*, pages 49–70. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, USA. xiii, 260.

diSessa, A. A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, 10:105–225.

diSessa, A. A. and Sherin, B. L. (1998). What changes in conceptual change. *International Journal of Science Education*, 20(10):1155–1191.

Dreyfus, T. and Tsamir, P. (2004). Ben's consolidation of knowledge structures about infinite sets. *Journal of Mathematical Behavior*, 23:271–300.

Fauconnier, G. and Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, New York.

Grady, J. (2005). Primary metaphors as inputs to conceptual integration. *Journal of Pragmatics*, 37:1595–1614.

Hammer, D. (1996). Misconceptions of p-prims, how might alternative perspectives of cognitive structures influence instructional perceptions and intentions? *Journal of the Learning Sciences*, 5:97–127.

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics*, 67 (Physics Education Research Supplement):S45–S50.

Hammer, D., Redish, E. F., Elby, A., and Scherr, R. E. (2004). Resources, framing, and transfer. In Mestre, J., editor, *Transfer of Learning: Research and Perspectives*. Information Age Publishing.

Hougaard, A. (2005). Conceptual disintegration and blending in interactional sequences: A discussion of new phenomena, processes vs. products, and methodology. *Journal of Pragmatics*, 37:1653–1685.

Mandler, J. M. (1992). How to build a baby: Ii. conceptual primitives. *Psychological Review*, 99(4):587–604.

McCloskey, M. (1983a). Intuitive physics. *Scientific American*, 248(4):122–130.

McCloskey, M. (1983b). Naive theories of motion. In Gentner, D. and Stevens, A. L., editors, *Mental Models*, pages 299–324. Lawrence Erlbaum Associates, Hillsdale, NJ.

Minsky, M. (1975). A framework for representing knowledge. In Winston, P., editor, *The Psychology of Computer Vision*. McGraw-Hill.

Minsky, M. (1985). *The Society of Mind*. Simon and Shuster, New York.

Sayre, E. C. and Wittmann, M. C. (2008). Plasticity of intermediate mechanics students**Y** coordinate system choice. *Phys. Rev. ST Phys. Educ. Res.*, 4(2):020105.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on process and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22:1–36.

Someren, W. M., Barnard, F. Y., and Sandberg, A. C. J. (1994). *The think aloud method. A practical guide to modleing cognitive process*. Harcourt Brace & Company, London.

Tsamir, P. and Dreyfus, T. (2002). Comparing infinite sets – a process of abstraction. the case of ben. *Journal of Mathematical Behavior*, 113:1–24.

Wandersee, J. H., Mintzes, J. J., and Novak, J. D. (1994). Research on alternative conceptions in science. In Gabel, D. L., editor, *Handbook on Research in Science Teaching and Learning*, pages 177–210. Macmillan, New York.

Wittmann, M. C. (2002). The object coordination class applied to wavepulses: Analysing student reasoning in wave physics. *International Journal of Science Education*, 24(1):97–118.

Wittmann, M. C., Steinberg, R. N., and Redish, E. F. (1999). Making sense of students making sense of mechanical waves. *The Physics Teacher*, 37:15–21.


## Acknowledgments

The work described in this paper was funded in part by NSF grants REC-0633951 and DUE-9455561. The data were gathered while the author was a graduate research assistant at the University of Maryland.

# Adapting Workflow Technology to Design-Based Research: Development of a Method for Organizing the "Messiness" of Research in Technology-Rich Online Learning Environments

Alan J. Hackbarth, Sharon J. Derry, Brendan R. Eagan, Julia Gressick, University of Wisconsin – Madison, Educational Sciences Building, 1025 W. Johnson St., Madison, WI 53706
E-mail: ajhackbarth@wisc.edu, derry@education.wisc.edu, beagan@wisc.edu, gressick@wisc.edu

**Abstract**: A fundamental challenge of design-based research is that there are many variables that affect success of a design. Designers collect large amounts of data, but limited time and resources make analysis difficult and conclusions uncertain. Workflow technology is utilized in business and applied science environments to automate work processes and reveal "know-how," often tacit in scientific processes, which facilitate multiple levels of reuse. We developed a method for representing activity in an experimental online course as workflow expressions. Bielaczyc's (2006) Social Infrastructure Framework (SIF) is utilized to identify important variables, and *comprehensive data mining* (CDM) techniques are used to recover data from course session modules and activity logs. In this paper we review the literature related to our theoretical framework, describe the CDM-based methodology, and give an example of how we are using it to support design-based research within an online college course.

## Introduction

This contribution reports a new design research methodology based on workflow technology (Georgakopoulos, Hornick, and Sheth, 1995; Ludäscher et al., 2005) and the Social Infrastructure Framework (SIF) (Bielaczyc, 2006) that is especially appropriate for online learning environments. The context for our work is HAL Online, an innovative experimental section of "Human Abilities and Learning," a required course for teacher education and other education-oriented majors offered through the Educational Psychology Department at the University of Wisconsin-Madison. HAL Online includes reading assignments, online exams, and meets in person occasionally, but the primary instructional method is online argumentative discourse in small group discussion forums. Small groups comprise 5-7 students who are grouped by the instructional staff based on their similar majors and career interests and who work as a team throughout the semester. There are about 5 groups and 30 students in each course offering. The broad goal for HAL Online students is to acquire foundational understanding of human learning and thinking that is grounded in current research and theory from cognitive science, socio-cultural theory, and neuroscience. Students must demonstrate such understanding by using course ideas to: A. construct and justify well-formed arguments about a range of important topics; B. articulate and evaluate the reasoning and arguments of others, including peers and school children; C. participate effectively in argumentative small-group discourse online; and D. work collaboratively to design, justify, and critique instruction that aims to promote good thinking in areas related to their disciplinary majors.

The course is divided into 4 units: Language and Reasoning; Brain-Based Education; The Mathematical Basis of Reasoning; and Instructional Design (a small-group project). Each unit comprises 3 or 4 interrelated modules, each lasting one week. There is a required face-to-face (FTF) class for each unit. In FTF classes students hear lectures, discuss readings, solve and discuss reasoning tasks with their group, and build community and rapport with their peers. In a typical week students are given a reading assignment from a text that they complete over a 2-day period. During this period they may also participate in a warmup activity online which often involves viewing a video. Following this individual preparation they are required to go online and participate with their small groups in a discussion activity. A typical discussion activity is "Brain Gym," part of the Brain-Based Education unit, in which students are asked to play the roles of parents and educators trying to decide whether either of two brain-based curriculum proposals should be offered by their school. Another is "Math Wars" in the mathematical reasoning unit, a discussion of whether a school should adopt a progressive mathematics curriculum and approach exemplified in videos viewed online.

In this paper we illustrate our proposed workflow methodology with "Adventures in Argument," a one-week module within the Language and Reasoning unit. In this activity students are required to observe and evaluate arguments made in the popular media about current issues in the news. For example, in one semester students evaluated arguments made in political debates. In a recent semester students evaluated arguments about different health reform bills. Students' participation in small-group discussions is graded in accordance with a rubric (see Table 1).

Table 1. Criteria for evaluating students' forum contributions

1. Do you make a sufficient number of contributions? There is no set limit or number required, but a good rule of thumb is *at least* 4 thoughtful posts per forum (not per discussion topic).
2. Are your arguments thoughtful, intelligent, mature, and justified with reasons and evidence (rather than just expressing personal opinion)?
3. Do your posts specifically connect the forum discussion topics to the readings, providing evidence that you are thoughtfully connecting ideas from the course to the forum issues?
4. Do you participate in a discourse (versus post at the last minute)? Forum discussions usually start on a day a little before the previous topic closes and they wrap up before the beginning of the next topic. Engaging in the forum discourse *throughout* the period rather than just throwing up a few posts at the end will improve your grade.
5. Have you been a good group citizen, taken on some leadership -- starting discussions, serving as chair or summarizer, helping keep the group on task, contributing positive and encouraging words to others?

At the end of each unit each student is required to submit a reflection on what they have experienced and learned in that unit. These submissions are "published" through the blog facility provided in the Moodle course management system (CMS) through which we offer the course online. At the end of each unit students individually complete an online quiz assessing their mastery of the material in the unit. A typical quiz question provides students with a link to a video and asks the students to analyze the video in some way. For example, in the mathematical reasoning unit, a video of a child solving a mathematical task is provided and students are asked to use course ideas to describe and analyze the child's reasoning and problem solving strategies.

## Design-Based Research (DBR)

Design-based research is grounded in the systematic design and study of instructional strategies and tools in authentic contexts (Barab and Squire, 2004); as such, there is no one accepted definition or methodology. The Design-Based Research Group (2003, p. 5) identifies characteristics of "good" design-based research. The critical characteristic is that the central goals of designing learning environments and developing theories or "proto-theories" of learning are intertwined. Development and research takes place through continuous cycles of design, enactment, analysis, and redesign, and research must account for how designs function in authentic settings – not only in terms of success or failure, but also in terms of understanding the learning issues involved.

One purpose of our research is to improve the effectiveness of our course through an iterative process in which we make deliberate design changes. Another purpose is to assess the effects of design changes that are made out of necessity, often due to constraints over which the instructional staff has little control. For example, when the course is taught in summer it must be squeezed into a four-week intensive time period, which is very different from the full semester course that is stretched out over a 16-week period. Because the primary instructional method is online argumentative discourse in small groups, a third purpose of our research is to propose, test, and develop "proto-theories" about how small groups in online environments collaborate to produce successful outcomes.

Collins, Joseph, and Bielaczyc (2004) describe challenges that a design-research team faces when implementing design experiments. Foremost is that research is usually conducted in the "blooming, buzzing confusion" of classroom learning environments (Brown, 1992). There are many variables that influence the success of a design, and many of those variables cannot be controlled (Collins, Joseph, & Bielaczyc, 2004, p. 19). Furthermore, each variable is part of a systemic whole; it is impossible to change one aspect of the system without creating perturbations in others (Brown, 1992). Yet it is important to identify the critical variables of a design and how they fit and work together in practice (Collins, Joseph, & Bielaczyc, 2004, p. 34). One needs a well developed profile of an implementation in order to analyze a design in terms of its key elements and their interactions, and to determine how exactly one design differs from another in ways that might impact outcome. Because of the number of variables to account for, design researchers usually end up collecting large amounts of data, more data than they have time or resources to analyze (Brown, 1992; Collins, Joseph, & Bielaczyc, 2004, p. 19). To ensure that design researchers maximize their use of the collected data, analytical procedures that organize and document data in an easily accessible format that facilitates more efficient, deeper analyses are needed.

Barron (2007, p. 178) discusses the emergence and value of *intermediate representations* as a response to this need. Focusing her discussion on video research, she argues that intermediate representations are important for identifying *what to analyze* and for *understanding patterns within and across* segments of video. Macro-level intermediate representations may also be the appropriate level for sharing the details of a successful designed intervention with practitioners. This would satisfy a core characteristic of DBR; that the goal of developing sharable, adaptable learning interventions of practical use to practitioners is intertwined with the goal of developing theories or "proto-theories" of learning (Collins, 1992). Inspired by a suggestion from Rutgers colleague Grace Agnew (personal communication, 2007), we explored the potential of workflow technology to provide such representations.

### Workflow Technology

The core features of workflow technology – organizing the data, tools, participants, activities, and flow of a system – are well suited to the challenges of design-based-research (DBR). DBR researchers must attend to a large number and range of variables and their interrelationships, and they must control or at least be aware of the variable manipulations that occur either intentionally or unintentionally, during each iteration of the design research process. Workflow technology is especially useful in online learning environments such as ours because many variable values and their relationships *can be recovered post hoc* from activity logs and other data created in the online course management system (CMS). Workflow expressions derived from such data can provide diagrammatic, annotated representations of a targeted part of a course as a system. Strauss

In the following sections we will review the literature that contributed to the theoretical framework for our workflow expressions approach and provide an example of how the visual models are used to support analysis in design-based research.

## Theoretical Framework

Core concepts of our workflow expressions model come from the Social Infrastructure Framework (Bielaczyc, 2006), the literature on business and applied science workflows (Georgakopoulos, Hornick, and Sheth, 1995; Ludäscher et al., 2005), and adaptation of a comprehensive data mining (CDM) model developed by the Collaborative Technology Research Group (CTRG) at the University of Colorado-Boulder (Rembert, 2006).

### The Social Infrastructure Framework (SIF)

The SIF identifies critical design elements to be aware of and accounted for when designing socio-techno learning environments, and articulates them in terms of four dimensions (Bielaczyc, 2006). These are summarized in Table 2:

Table 2. Critical design elements of socio-techno learning environments.

| Dimension | Characteristics |
|---|---|
| Cultural beliefs | Cultural beliefs are not *designed* per se but cultivated over time, and influence such things as: how learning and knowledge are conceptualized, goals, how identities of students are shaped, how the identity of the teacher is understood, and how technology use is viewed. |
| Practices | Concerns the ways in which teachers and students engage in activities with tools, how students work on activities, how groupings of students are organized, what roles students play within groups, modes of interaction supported or constrained within and between groupings, and the role the teacher plays. |
| Socio-techno-spatial relations | Concerns the organization of physical space and technology workspaces as they relate to teacher and student interactions. Considerations include the physical organization of classrooms, how groups incorporate technology, the affordances and constraints of the technology, how pervasive technology use is in the intervention, how access is provided |
| Interactions with the "outside world" | Refers to the online and offline ways students are able to interact with people and be influenced by their engagement in events outside their immediate classroom environment. Aspects to consider include issues of authenticity: How is knowledge brought in from and extended to the outside world? |

Consideration of what variables can capture the characteristics of these dimensions informs the design of data collection from online modules, highlights what should be mined from available activity logs, and influences the look of a resulting workflow expression by identifying what information needs to be represented

## Workflows in Business and Science

*Workflow* is a concept closely related to the design, or reengineering, of business and information processes (Ludäscher et al., 2005). In a broad and multi-leveled sense a workflow expression is the diagrammatic result of "capturing" the workings of a system from beginning to end.

Within the literature two broad versions of a standard definition of workflow emerge: business and scientific. Georgakopoulos, Hornick, and Sheth, (1995) describe a *business-oriented* workflow as a collection of tasks organized to accomplish some business process, performed by one or more software systems, one or a team of humans, or a combination of these. A workflow defines the order of task invocation, task synchronization, and dataflow. S*cientific workflows* are valuable knowledge assets in their own right because they are graphical representations of scientific "know-how" that is often tacit. A key feature in the development of scientific workflows is the notion of reuse (De Roure, Goble, and Stevens, 2008). Reuse can occur at multiple levels. With different parameters and data, fragments and patterns of workflows can be reused to support science outside their initial application, or they can provide a means of codifying, sharing, and spreading the workflow designer's practice.

Both business and scientific workflow expressions attempt to define processes of a system in terms of the tasks performed, order, resources consumed and produced, and relationships between person and machine. Both account for control of events and flow of data. However, when analyzing the underlying principles of the approaches, Ludäscher et al. (2005) found a focus on *control flow patterns* and *events* – what was done, and in what order – in business-oriented systems, while scientific workflow systems tended to have execution models that were more *dataflow-oriented* – interest in how data gets passed through, transformed, and used by the system (p. 1046).

The concepts of reuse, control flow, and data flow are all important to account for in an educational workflow model; they facilitate sharing and analysis of adaptable interventions. However, the unpredictability of educational interventions noted by Brown (1992) suggests that educational workflows cannot be preprogrammed in the way that many business or scientific-oriented workflows are (e.g., loan applications or mineral identification systems). The important functions of educational workflows, then, include capturing and representing: 1) the intended, adaptable aspects of interventions that can be shared and reused across diverse educational and research settings; and 2) the actual and detailed data flow within interventions, which supports scientific analysis. This suggests two kinds of educational workflows; the former, which we will call *macro-flows*, may be roughly equated with educational *macro-scripts*– general pedagogical models, often representing intentions, that aim to create learning situations in which productive interactions and outcomes will hopefully occur (Dillenbourg and Hong, 2008). The latter are *micro-flows*, which are like "micro-scripts," although they may emerge during practice and are determined post-hoc. Micro-flows represent more detailed models of actual individual and group activity. In online environments, data for both types of workflow representations can be mined post hoc from activity logs and other data found in the course management system (CMS). We hypothesized that both would be useful in design-based research.

## Comprehensive Workflow Mining

Much of the literature on workflow modeling focuses on the use of Petri-Nets to automate and analyze business processes (e.g. van der Aalst, and van Hee, 2004), a highly structured, mathematics-based methodology that describes processes in terms of weighted nodal relationships among places and transitions. This workflow method's focus on control flow makes it too narrow to describe the relevant dimensions of a learning environment. However, the Collaboration Technology Research Group (CTRG) at the University of Colorado-Boulder has developed a comprehensive workflow modeling language called Information Control Nets (ICN) that is graphical and intuitive and broadens the scope of workflow mining to include a wider range of perspectives (Rembert, 2006). Briefly, the primary perspectives focused on by this modeling methodology include: the **functional** – *what* tasks or activity takes place; the **control flow** – *when* tasks are done; the **informational** – *which* data are processed and the *data flow* of the process; the **resource** or **organizational** – *who* or *what* performs a task; and the **operational** or **application** perspective – *how* a task gets done.

The ICN workflow modeling language has a mathematical and graphical representation; for the purpose of the present paper we focus only on the graphical elements of representation. These elements are shown and described in Table 3:

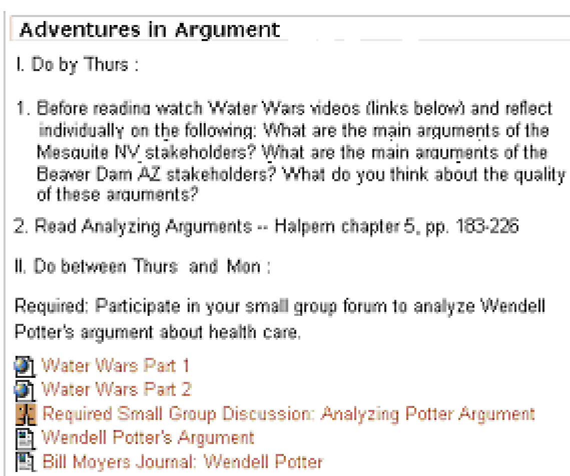
Table 3. Graphical elements of the ICN workflow modeling language.

| | | | |
|---|---|---|---|
| | *Tasks* (represented by labeled circles) are a unit of work carried out by one or more people. | | *Data repositories* (represented by labeled squares) have production/consumption relationships with tasks - they provide data that is consumed by a task or collect and hold data produced by a task. |
| | *Roles* (rounded rectangles) can represent individuals or groups, specified in terms of features such as responsibilities, authority, and availability. | | *Participants* are represented as labeled stick figures. An organizational population specifies which participants belong to which role/group, as well as relationships amongst participants. |
| | If a task requires a certain *application type* (hexagon), to be completed, then that task might be accomplished by any *application instance* (monitor). | | There are four *control tasks*; two branch tasks – parallel (closed circle) and conditional (open), and two join tasks – synchronization (closed circle) and merge (open). |

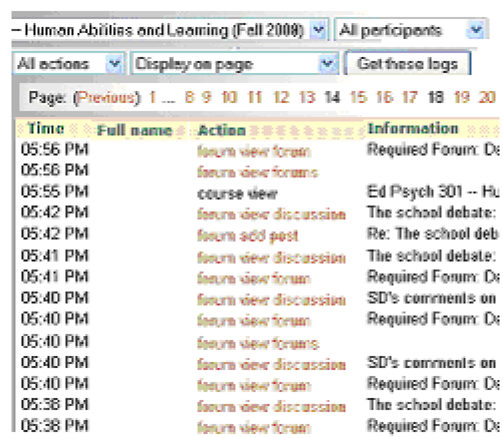
## Example of Analysis using CDM-based Workflow Expressions

### Mining Data and Constructing Workflow Expressions from Adventures in Argument

The Moodle session module shown in Figure 1(a) provides information to students about what activities to do, in what order, and provides links to online resources and tools that are needed for activities. (The link to the discussion provides further instructions to students, and hence workflow information, that is not shown here).




(a)




(b)

Figure 1. HAL Online module seen by students (left) and partial activity log (right).

Such information is used to construct the *macro-workflow* expression shown in Figure 2. Each activity is represented in a circle and the flow from one activity to the next is shown with arrows. Closed dots before the 'View Video' activities indicate that both activities must be done, but order doesn't matter. Open dots after indicate that activity can proceed even if a student only views, for example, one of the two videos. Percentages by each circle indicate the level of compliance for two groups that were examined in this example. These data were "discovered" by mining the activity logs (see Figure 1(b)). Information about tools, roles and data repositories are represented using ICN graphical elements (Table 3). We added new icons for special data repositories unique to educational work: a book for textbook-type information, and a silhouette for representing students' activated prior knowledge.

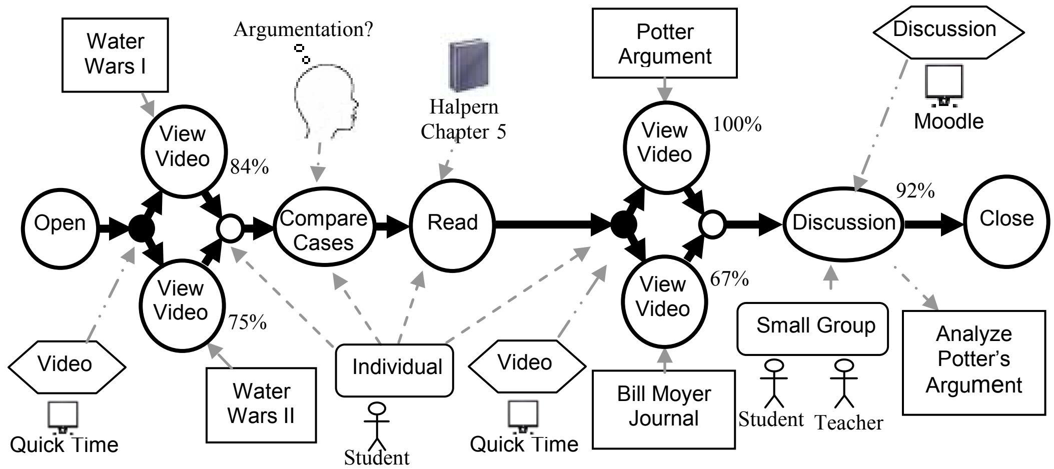



Figure 2. Macro-workflow expression for "Adventures in Argument".

Figure 3 illustrates how logs specific to Moodle tools (i.e., the discussion tool) and resources (i.e., videos) were used to construct *micro-workflows* for the discussion activity (part of the macro-flow) for students in two contrasting groups in our course. Micro-flows were organized in two musical-type "scores" above and below a master timeline that allowed us to see the flow of data for individual students in each group as they engaged in the discussion activity. In the diagrams, small open squares indicate when students viewed videos. The numbers inside the circles indicate thread number (left) and post-in-thread (right). Closed squares on the circles indicate that the post connects to a video or the assigned reading (which one can be seen specifically in the discussion log).

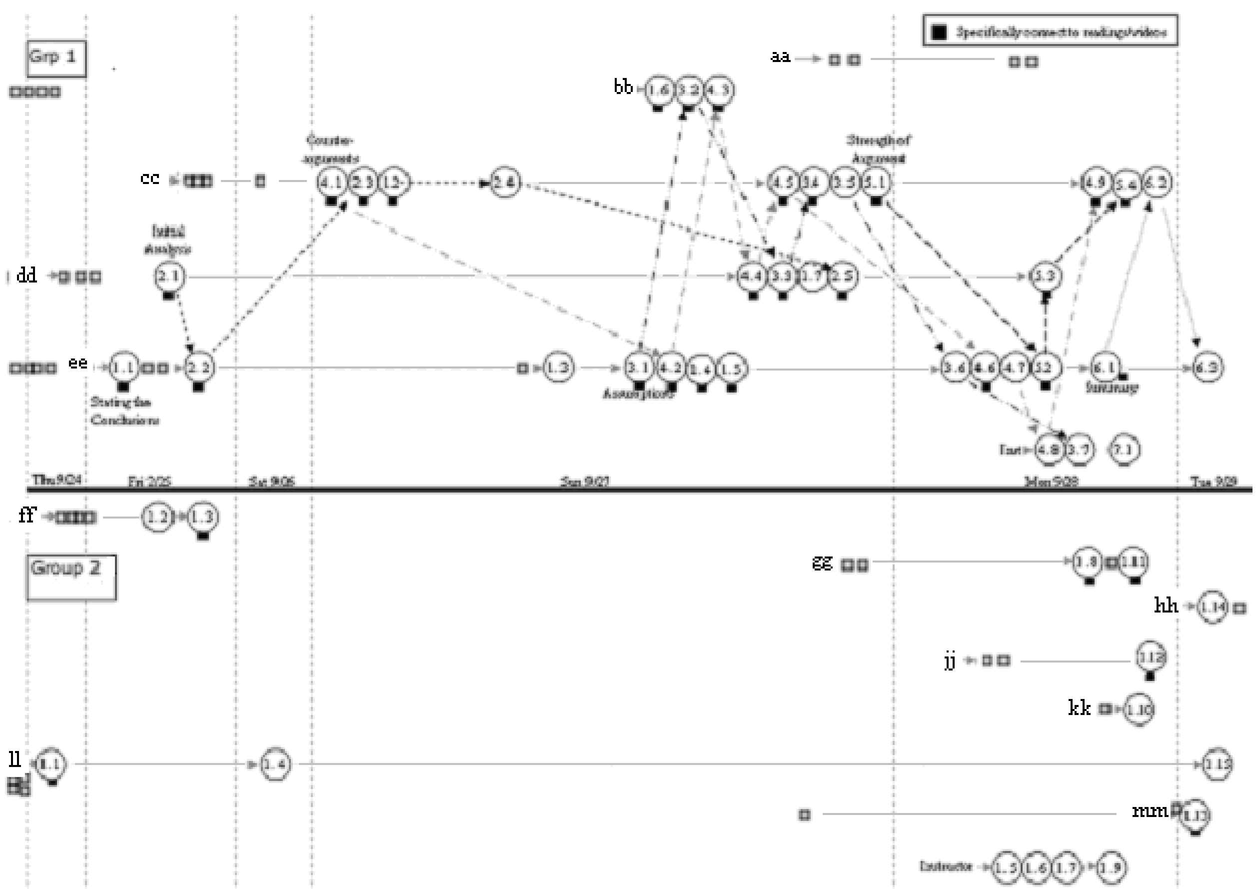



Figure 3. "Adventures in Arguments" micro-workflow expression for forum discussion of two groups.

## Analysis

When analyzed by the instructional development team, both the macro- and micro-flow representations helped the research team generate many ideas regarding how well the Adventures in Argument module worked and in suggesting design changes for the next iteration. Even a cursory glance at the micro-workflows for the two groups (Figure 3) revealed stark differences in two groups. Group 1 generated a greater number of threads, postings, made more references to course readings and videos, and referenced one another's postings. Three different members started threads over the course of the discussion. By comparison, Group 2's postings were contained in one thread, postings were disconnected, no student posted more than three times, and four members made only one post. A closer examination shows that most members of Group 1 accessed *all* of the supporting videos early and often and even re-viewed certain videos during the discussion forum. In contrast, only members of Group 2 viewed all four videos before beginning the forum discussion; the majority of members viewed *some* videos right before making their posting. (In light of this, the compliance statistics in the macro-workflow are a little misleading; Group 1 actually had 100% compliance viewing the videos where Group 2's overall compliance was less than average for the class.) As well as Group 1 appears to do in comparison to Group 2, there is room for improvement. One member (OB) viewed the videos very late in the work process and did not participate in the discussion. In her reflection on the experience with the unit she reported that she had trouble viewing the videos and felt she was too late to make a contribution to the discussion by the time she fixed the problem.

The interpretive power of workflows is illustrated in reflections by the professor-researcher in charge of the course, who made the following observation:

> "I noticed some things that are possibly important. One thing was the similarity of Group 1 to a strong group in a study of five online groups that Mary and I did awhile back. The strongest group in that study had a strong leader who organized the discussion space, breaking the discussion down into sub-topics, starting some topics in a certain direction, sometimes giving particular assignments. In Group 1, CC takes a leadership role and does something similar that turns out to be important: He wrote *'I would propose that we start a separate thread for each of the main areas of analyzing the argument, such as conclusion(s), premises and omitted premises, counterarguments, assumptions, etc. I have started one on Conclusions. This organization will help us get a lot of material in an organized way which will hopefully make it easier to summarize in a couple of days.'* Also, Group 1 was far from perfect; there were non-participants. Students who were left out expressed frustration in their blogs with their inability to interact with the technology. These "weaker" students seem intellectually capable, but the technology may be putting them on overload."

Additional analysis of the macro-workflow expression (Figure 2) revealed that there were two activities – comparing cases and reading the text chapter – for which we could not verify compliance. Both activities relied to some extent on data repositories that were located outside the Moodle logging system. Students' compliance with reading of the text could be inferred from a closer examination of the students' postings, but the comparing cases activity, which had a theoretical rationale, could not be evaluated due to insufficient data. One role of the workflow expressions in DBR is to focus attention on design flaws that exist because an activity fails to generate log data that can be mined.

## Conclusions

Our team was able to use the macro- and micro-flow expressions to guide us in making design changes to be implemented in the next iteration of our course. Examples under consideration include: 1) Create more explicit scaffolding on collaboration to be incorporated into the instructions for the discussion task – i.e., choose a leader, break discussion down into sub-topics, start discussion threads for each sub-topic, distribute tasks; 2) Emphasize in the discussion-task instructions that a post needs to include explicit references/connections to videos, readings, and/or other postings; 3) Mandate that students access resources early in the session (perhaps tying a grade to *when* they access or comment on resources). At the process level we decided to include logable tasks for all activities, e.g., write a brief summary of the case comparison in the warmup activity, or complete a short questionnaire about a reading.

## Discussion

The example above illustrates how workflow expressions derived from data mined from the online session modules and activity logs of an online course can help facilitate the analysis of a complex environment by presenting visual representations that incorporate design variables of interest.

Both workflow expressions served a valuable purpose. The micro-workflow expression allowed us to see when and how individual students engaged in activities – their *control flow* – and how group members interacted with one another during collaborative activities, how they used course resources, and how they took up and used data created by one another – their *data flow*. The micro-workflow expressions facilitated hypothesis generation about success in group argumentative activities – that students need to organize their discussion space by breaking down arguments into sub-topics, starting some topics in a certain direction, and distributing tasks. This hypothesis amounts to a proto-theory about successful online argument as pedagogy, supporting a dual focus on theory-building and instructional design. Finally, micro-workflow expressions help identify problems that may be related to course tools or processes, e.g., inability to access videos. These kinds of insights also lead to instructional design changes.

Macro-workflow expressions provide a means for researchers to see all design elements of an intervention and their relationships to one another. This is important because it allows researchers some sense of control as they contemplate design changes. As Ann Brown (1992) pointed out, it is impossible to change one aspect of the system without creating perturbations in others. Macro-workflow expressions allow the researcher to focus design changes on specific areas of macro-flow while striving to monitor activities and achieve some level of control within others. An important goal of design-based research is to continually adapt an instructional design given new insights about the design, but an equally important question to ask is, "How much adaptation is taking place?" We believe the workflow expressions derived from course session modules and activity logs has been a valuable asset as we strive to answer this question.

## References


Barab, S.A., & Squire, K. (2004). Design-based research: Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*, 1-14.

Barron, B. (2007). Video as a tool to advance understanding of learning and development in peer, family and other informal learning contexts. In R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.) *Video research in the learning sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bielaczyc, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with technology. *The Journal of the Learning Sciences*, *15*(3), 301-329.

Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141-178.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, *13*(1), 15-42.

De Roure, D., Goble, C., & Stevens, R. (2009) The design and realization of the myexperiment virtual research environment for social sharing of workflows. *Future Generation Computer Systems*, *25*(5), 561-567.

Dillenbourg, P. & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning*, *3*, 5-23.

Georgakopoulos, D., Hornick, M., and Sheth, A. (1995). An overview of workflow management: From process modeling to workflow automation infrastructure. *Distributed and Parallel Databases*, *3*, 119-153.

Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, E.A., Tao, J., and Zao, Y. (2005). Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience*, *18*(10), 1039-1065.

Rembert, A.J. (2006). Comprehensive workflow mining. *ACM Southeast Regional Conference: Proceedings of the 44th Annual Southeast Regional Conference*, March 10-12, Melbourne, FL.

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5-8.

van der Aalst, W.M.P., & van Hee, K. (2004). *Workflow management: Models, methods and systems.* Cooperative Information Systems Series. MIT Press.


## Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 0822189. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

# Coordinating Collaborative Problem-solving Processes by Providing Part-task Congruent Representations

**Abstract:** This study investigated whether structuring an online collaborative problem-solving task into part-tasks and providing part-task congruent representations supports students establishing and maintaining a shared understanding of the concepts, principles and procedures in the knowledge domain and negotiating about them. This better coordination of students' discussions was, in turn, expected to lead to better problem-solving performance. In triads secondary education students worked on a case-based business-economics problem in four conditions. In one condition, groups received three representations, each congruent to one of the three part-tasks. In the other three conditions, groups received one of the representations for all three part-tasks, thus a representation congruent to one part-task, but incongruent to the other two. The results show that coordination processes were indeed more suited. That is, students where better able to establish and maintain a shared understanding and to negotiate about it and, therefore, led to better problem-solving performance.

## Introduction

Solving a complex problem is regarded as a sequenced phased process (i.e., problem orientation, problem solution, solution evaluation) in which each phase has its own specific purpose and where each phase requires a specific kind of interaction (Ploetzner, Fehse, Kneser, & Spada, 1999; Van Bruggen, Boshuizen, & Kirschner, 2003). Coping with the task demands of the different problem phases (i.e., part-tasks), requires students to actively engage in a process of sense-making of the knowledge domain by articulating and discussing multiple problem perspectives and problem-solving strategies. That is, students are required to carry out part-task related activities such as discussing the concepts, principles, and procedures of the domain (Barron, 2003; Hmelo-Silver, Duncan, & Chinn, 2007). Meaningful discussions about the domain can, however, hardly be achieved when students are not aware of each others' knowledge, ideas and do not negotiate about them with their peers. Students are required to activate their knowledge and skills and to establish and maintain a shared understanding of the domain; a common frame of reference where conflicting points of view be detected and negotiated. In this way collaborative learning situations require three main processes of coordination: (1) mutual activation and sharing of knowledge and skills, (2) grounding or creating a common frame of reference, and (3) negotiation or the process of coming to agreement (Erkens, Jaspers, Prangsma, & Kanselaar, 2005; Kirschner, Beers, Boshuizen, & Gijselaers, 2008). Research on Computer Supported Collaborative Learning (CSCL) has shown that representational tools can beneficially affect student interaction by providing external representations (see Fisher, Bruhn, Gräsel, & Mandl, 2002). First, an ER offers a restricted view of the knowledge domain, guiding the content-related interaction in a specific manner (Suthers, 2006). Second, in their discussions students can refer to the ER (i.e., deictic referencing), thereby supporting them to create a common frame of reference and facilitating a meaningful discussion (Suthers, Hundhausen, & Girardeau, 2003). These studies, though very valuable and informative, often neglect the fact that problem-solving tasks are usually composed of fundamentally different part-tasks that each require a different perspective on the knowledge domain. Therefore, multiple representational tools each containing a different external representation seem to be required. To be supportive for problem-solving, the external representation provided in the representational tool has to be matched to the part-task demands and activities of a specific problem phase. Otherwise, communication problems can occur and problem-solving performance might decrease (Van Bruggen, et al.).

The goal of the study presented in this paper is to twofold. On the one hand it is aimed at determining whether proper coordination processes can be evoked through providing part-task congruent guidance in the representational tools. On the other hand it is aimed at determining whether such an approach can lead to better problem-solving performance.

## Coordination Processes

For meaningful discussion to arise, students must properly coordinate their discussions of the concepts, principles and procedures by carrying out communicative activities such as making their own knowledge and ideas explicit to other group members, focusing, checking and argumentation (Andriessen, Baker, & Suthers, 2003; Erkens, et al., 2005; Weinberger Ertl, Fischer, & Mandl, 2005). When made explicit, students must try to maintain a shared topic of discourse and to repair a common focus if they notice a focus divergence. Students coordinate their topic of discourse by *focusing*. Also, not all concepts, principles, and procedures are relevant for carrying out a part-task, thus, students have to guard the coherence and consistency of their shared understanding of the knowledge domain. By *checking*, students ground their communication in a common understanding, which was found to be one of the major communicative activities in dialogues of collaborative problem-solving and related to the quality of the problem solving process (Van der Linden, Erkens, Schmidt, & Renshaw, 2000). Furthermore, students must come to agreement with respect to relevant concepts principles and procedures. By *argumentation* they will try to change their partners viewpoint to arrive at the best way to carry out a part-task or at a definition of concepts acceptable for all. In this argumentation process they try to convince the other(s) by elaborating on their point of view, giving explanations, justifications and accounts (Erkens, et al.; Kirschner, et al., 2008). Only when students carry out such communicative activities their interaction can be sufficiently coordinated and multiple perspectives on the problem and the problem-solving strategy can arise.

## Providing Part-task Congruent Representations

Due to its different part-tasks, problem-solving tasks require multiple representational tools providing different external representations. The specific ontology (i.e., objects, relations, rules for combining them) of each ER offers a restricted view of the knowledge domain (i.e., problem representations), guiding student interaction in a specific manner. To effectively do this, one must carefully match the external representation provided in a representational tool to the part-task demands and activities of a specific problem phase (Van Bruggen, et al., 2003). To ensure this alignment between tool, tool use and part-tasks scripting is employed (Dillenbourg, 2002; Weinberger, et al., 2005). According to Dillenbourg a script is “a set of instructions regarding to how the group members should interact, how they should collaborate and how they should solve the problem” (p. 64). Integrating scripting with the availability of representational tools sequences and makes the different part-task demands explicit so that they can be foreseen with part-task congruent guidance in the representational tools (see Table 1). By doing so, communicative activities beneficial for coordinating the collaborative problem-solving can be evoked.

Table 1: Congruence between external representation and part-task demands

| Problem phase | Part-task demands | ER | Representational guidance |
|---|---|---|---|
| Problem orientation | Determining core concepts and relating them to the problem | Conceptual (static) | Showing concepts and their interrelationship |
| Problem solution | Proposing multiple solutions to the problem | Causal (static) | Showing causal relation between the concepts and possible solutions |
| Solution evaluation | Determining suitability of the solutions and coming to a final solution to the problem | Simulation (dynamic) | Showing mathematical relation betweenthe concepts and enabling manipulation of their value |

## Design and Expectations

This study focuses on whether providing part-task congruent support in the representational tools affects both students’ communicative activities and problem-solving performance in a CSCL-environment. In four experimental conditions, student triads had to collaboratively solve a case-based problem in business-economics that was divided into three problem phases each coupled with a different ER. To study the effect of condition, ERs were either matched or mismatched to the different problem phases; in other words they were either congruent or incongruent to the required task activities. In three mismatch conditions, groups received either a static ER (i.e., conceptual or causal ER) or a dynamic ER (i.e., simulation) that matched only one of the part-tasks. The scripting structured the problem-solving process in three phases, but only one of the ERs is available to the students for solving the problem yielding a phase-match when the ER matched one of the three phases and a mismatch for the other two phases. In the fourth condition, groups received all three ERs in a phased order receiving the ER most suited to each problem phase. Here, thus, there was a match between all three ERs and all three part-tasks. Due to the presumed match between ERs and part-tasks, student understanding and communicative activities were expected to increase, allowing the students to reach better problem solutions. It was, therefore, hypothesized that students in the match condition *(H1)* carry out more communicative activities beneficial for coordinating their collaborative problem-solving and *(H2)* have better problem-solving performance. In this paper, the focus will be on guiding student interaction when collaboratively solving a complex business economics problem. One should, therefore, take into account that the congruency of the content-related guidance was tailored to the part-task demands of this problem. The premise behind the design of the representational scripting and its use, however, can be generalized to all situations where a problem-solving task has a part-task structure.

# Method

## Participants

Participants were students from a business-economics class in a secondary education school in the Netherlands. The total sample consisted of 39 students (24 male, 15 female). The mean age of the students was 16.74 years (*SD* = .83, Min = 15, Max = 18). The students were randomly assigned to 13 triads divided between the four conditions; four triads in the conceptual and three triads in each of the other conditions (i.e., causal, simulation and match).

## Problem-solving task and materials

### CSCL-environment: Virtual Collaborative Research Institute

Students worked in a CSCL-environment called Virtual Collaborative Research Institute (VCRI, see Figure 1), a groupware application for supporting the collaborative performance of problem-solving tasks and research projects (Broeken, Jaspers, & Erkens, 2006). For this study, five VCRI tools were augmented with representational scripting. All tools, except the Notes tool, were shared among group members.

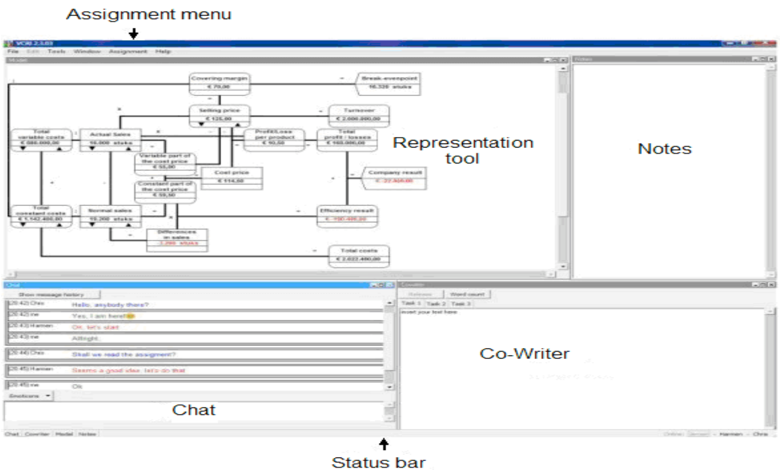



Figure 1. Screenshot of the VCRI-program.

The *chat tool* enables synchronous communication and supports students in externalizing and discussing their knowledge and ideas. The chat history is automatically stored and can be re-read by the students. In the *Assignment menu*, students can find the description of the problem-solving task / part-tasks. Besides this, additional information sources such as a definition list, formula list, and clues for solving the problem were also available here. The *Co-writer* is a shared text-processor where students can collaboratively formulate and revise their answers to the part-tasks. The *Notes tool* is an individual notepad that allows students to store information and structure their own knowledge and ideas before making them explicit. The *Status bar* is an awareness tool that displays which group members are logged into the system and which tool a group member is currently using. Students in all conditions had access to these tools and information sources. In other words, the different conditions were information equivalent and only differed in the way that the ERs are intended to guide the interaction.

### Problem-solving task and part-task congruent representations

All groups worked on a case-based problem in business-economics in which they had to advise an entrepreneur about changing the business strategy to increase profits (i.e., company result). To come up with an advice, students had to carry out three different part-tasks, namely (1) determine the main factors affecting the company's result and relate them to the problem, (2) determine how certain interventions affect company result, and (3) compare the effects of these interventions and formulate a final advice based on this comparison. Through scripting, the problem-solving process was structured into a problem orientation phase, problem solution phase, solution evaluation phase each focusing on one of the part-tasks. All groups were 'forced' to carry out the part-tasks in a predefined order; they could only start with a new part-task after finishing an earlier part-task. When group members agreed that a part-task was completed, they had to 'close' that phase in the assignment menu. This 'opened' a new phase, which had two consequences for all groups, namely they (1) received a new part-task, and (2) had to enter their new answers in a different window of the Co-writer. All conditions received the part-tasks in the same order, but only groups in the match condition received a new ER.

The *problem orientation phase* focused on creating a global qualitative problem-representation by asking students to explain what they thought the problem was and to describe what the most important factors were that influenced the problem. During this phase, students received the conceptual ER (i.e., a static content scheme; see Figure 2) that made two aspects salient, namely the core concepts needed to carry out this part-task and, how the core concepts are qualitatively interrelated. Students could see that 'company result' is affected by the 'total profit' and the 'efficiency result'. Such information should make it easier for them to create an overview of all relevant concepts, supporting them in finding multiple solutions to the problem in the following phase.

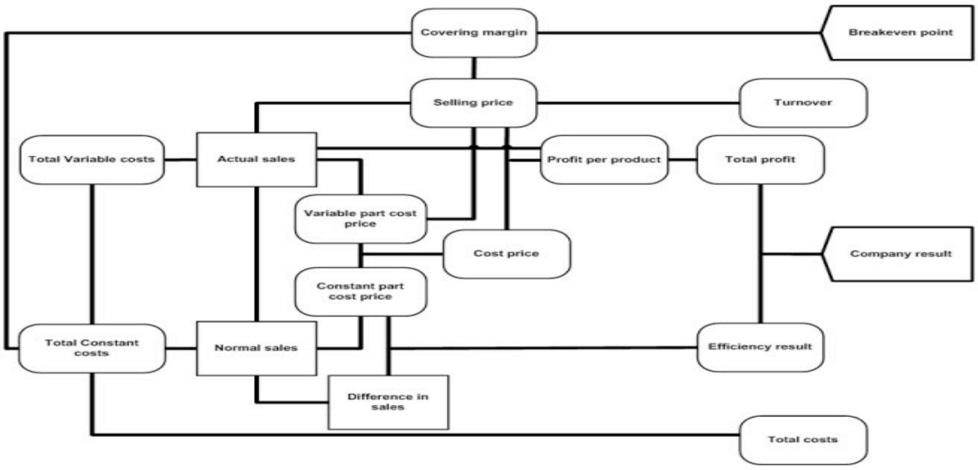



Figure 2. Conceptual ER.

The *problem solution* phase focused on creating a causal problem representation (i.e., making underlying business-economics principles explicit) by asking students to formulate several solutions to the problem. During this part-task, students received the causal ER (i.e., a static content scheme; see Figure 3), in which the causal relationships - visible through the arrows showing direction of the relationship between the concepts - were specified. The causal ER also contributed to increasing students' qualitative understanding by providing the students with possible interventions (i.e., changes in the business strategy), that each had a different effect on the company results. This should make it easier to effectively explore the solution space and should, in turn, support students in finding multiple solutions to the problem. Students could, for example, see that a PR-campaign affects 'actual sales that in turn affects 'total profit'.

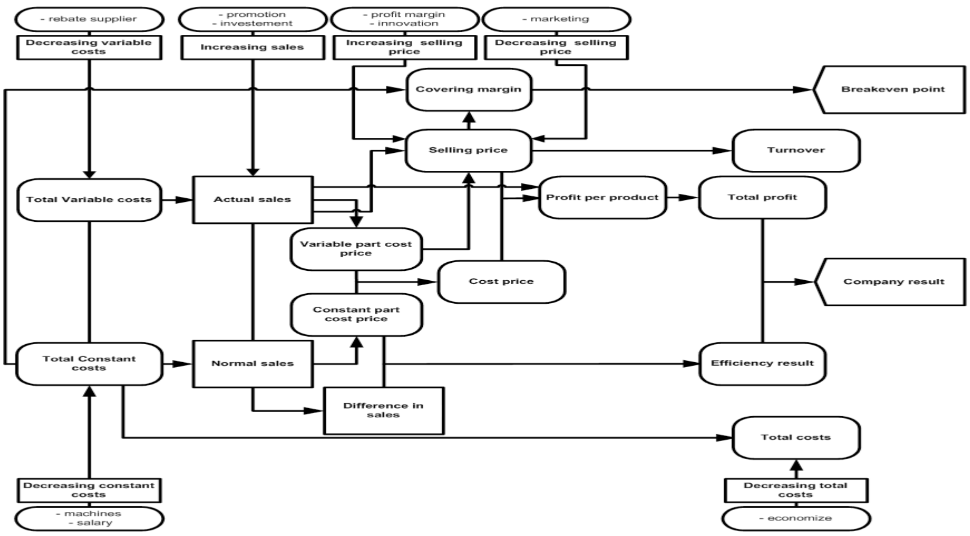



Figure 3. Causal ER.

Finally, the *solution evaluation phase* focused on increasing the students' understanding of the knowledge domain with the aid of a quantitative problem representation. Students were asked to determine the financial consequences of their proposed solutions and to formulate a final advice for the entrepreneur by negotiating the suitability of the different solutions with each other. During this phase, students received a simulation ER (i.e., a dynamic content scheme; see Figure 4) that enabled them to manipulate the values of the concepts by clicking on the arrows in the boxes. When the value of a certain concept was changed (i.e., increased or decreased), the simulation model automatically computed the values of all other concepts. This is meant to facilitate the determination and negotiation of the suitability of the different proposed solutions and reaching a final advice. Students could, for example, test how the PR-campaign affects the 'actual sales' and whether this in turn affects the 'total profit'. Only the simulation ER is capable of providing this kind of support, because the relationships between the concepts in this ER were specified as equations.

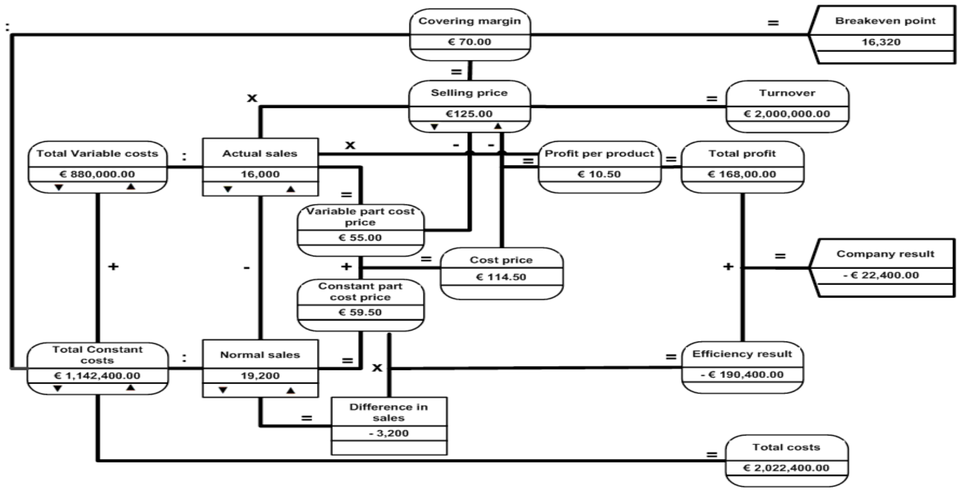



Figure 4. Simulation ER

## Procedure

All groups spent three, 70-minute, lessons solving the problem during which each student worked on a separate computer in a computer classroom. Before the first lesson, students received an instruction about the CSCL-environment, the group composition, and the problem-solving task. Students worked on the problem in the computer classroom, where all actions and answers to the part-tasks were logged.

## Measures

### Problem-solving performance

To measure the effect of condition on problem-solving performance an assessment form for all criteria of the problem-solving task was developed (see Table 2). All 41 items were coded as; 0, 1 or 2, whereby a '2' was coded when the answer given was of high quality (e.g., more suitable). Groups could, thus, achieve a maximum score of 82 points (41 * 2 points).

Table 2: Problem solving performance; items and reliability

| Criteria | Description | Items | α |
|---|---|---|---|
| Suitability | Whether the groups' answers were suited to the different part-tasks | 9 | .81 |
| Elaboration | Number of different business-economics concepts or financial consequences incorporated in the answers to the different part-tasks | 9 | .56 |
| Justification | Whether the groups justified their answers to the different part-tasks | 9 | .71 |
| Correctness | Whether the groups used the business-economics concepts and their interrelationships correctly in their answers to the different part-tasks | 9 | .68 |
| Continuity | Whether the groups made proper use of the answers from a prior problem phase | 2 | .67 |
| Quality advice | Whether the groups gave a proper final advice<br>- Number of business-economics concepts incorporated in the advice<br>- Number of financial consequences incorporated in the advice<br>- Whether the final answer conformed to the guidelines provided | 3 | .76 |
| Total score | Overall score on the problem-solving performance | 41 | .92 |

## Coordination processes

To examine the effect of condition data concerning students' coordination processes was collected by logging the chat-utterances of the group members. The content of these chat-protocols is assumed to represent what students know and consider important for carrying out their problem-solving task (Chi, 1997). A dialogue act is regarded as a communicative action that is elicited for a specific purpose representing a specific function in the dialogue (Erkens, et al., 2005), namely:

- argumentatives; indicating a line of reasoning,
- responsives; indicating responses to questions or proposals,
- informatives; indicating a transfer of information, often statements or evaluations,
- elicitatives; indicating questions or proposals requiring an answer,
- imperatives; indicating commands to take action or to draw attention.

Dialog-act coding was based on the occurrence of characteristic words or phrases (i.e., discourse markers) that indicated the communicative function of an utterance. The chat-protocols were searched for the occurrence of these discourse markers that led to the identification and coding of the dependent variables (see Table 2). This was automatically done with a MEPA-filter using 'if-then' decision rules that uses pattern matching to find typical words or phrases. Reliability of the dialogue act coding is 79% (Erkens & Janssen, 2008). After coding, score-frequencies for each dialogue act were computed and combined resulting in the measurement of the dependent variables.

Table 3: Coding of Students' Communicative Activities

| Activities | Dialogue Act | Description | Example discourse marker |
|---|---|---|---|
| Focusing | Elicitative proposal for action | Proposition for action | Shall we get started with the first part-task? |
| | Elicitative question open | Open question with a lot of alternatives | What do you think we should do next? |
| | Imperative action | Command to perform an action | Write the conclusion |
| | Imperative focus | Command for attention | Look at the representational tool |
| | Elicitative question verify | Question that can only be answered with yes or no | Do you refer to the company result? |
| Checking | Elicitative question set | Question where the alternatives are already given (set) | Are you for / against increasing sales? |
| | Responsive confirm | Confirming answer | Yes, sure |
| | Responsive deny | Denying answer | No, not |
| | Responsive accept | Accepting answer | Oh, Yes |
| Argumentation | Argumentative reason | Reason | Because…. |
| | Argumentative against | Objection | But … |
| | Argumentative conditional | Condition | If … (then …) |
| | Argumentative then | Consequence | Then … |
| | Argumentative disjunctive | Disjunctive | …or… or |
| | Argumentative conclusion | Conclusion | Thus… |

## *Data analyses*

In CSCL, group members influence each other (i.e., behave more or less similarly) causing non-independence of measurement (Kenny, Kashy, & Cook, 2006). This is problematic because many statistical techniques assume score independence and a violation compromises interpretation of the analyses (see Kenny, et al.). Non-independence was determined by computing the intraclass correlation coefficient and its significance (Kenny, et al.), for all dependent variables for student interaction. This resulted in non-independence ($\alpha < .05$) for all tests, justifying *Multilevel analysis* (MLA). MLA compares the deviance of an empty model and a model with one or more predictor variables to compute a possible decrease in deviance. The latter model is considered a better model when there is a significant decrease in deviance compared to the empty model (tested with a $\chi^2$-test). All reported $\chi^2$-values were significant ($\alpha < .05$) and, therefore, the estimated parameter of the predictor variables (i.e., effects of condition) were tested for significance. Due to the detection of outliers, the utterances of some students were deleted from the MLAs.

# Results

## Coordination processes

MLAs revealed that experimental condition significantly predicted the management of the interaction in the content space. The mean scores, standard deviations and the main effects (i.e., difference between match condition and non-matched conditions) for the communicative activities are listed in Table 4. A main effect for *coordination* was found, namely that students in the match condition, in total, exhibited more communicative activities compared to students in three non-matched conditions ($\beta = 37.36$, $p = .01$). For the specific communicative activities, the following results were obtained. First, a main effect was found for *focusing*; students in the match condition verified whether they were discussing the same topic and repaired the common focus more often than students in the non-matched conditions ($\beta = 8.48$, $p = .01$). When comparing the match condition to the other conditions separately, the effects were significant between all conditions. Second, a main effect for *checking* was found; students in the match condition devoted more attention to guarding the coherence and consistency of their shared understanding of the content space than students in the non-matched conditions ($\beta = 16.37$, $p = .02$). When comparing the match condition to the other conditions separately, this was the case for both the conceptual ($\beta = 16.02$, $p = .00$) and the simulation ($\beta = 23.20$, $p = .00$) conditions. Finally, a main effect was found for *argumentation*; students in the match condition exhibited more argumentative activities than students in the non-matched conditions ($\beta = 12.19$, $p = .00$). When comparing the match condition to the other conditions separately, this was the case for the causal (14.19, $p = .03$) and simulation ($\beta = 15.84$, $p = .04$) conditions.

As expected, students in the match condition were better able to establish and maintain shared understanding of the content space and negotiate about it better than students in the non-matched conditions. This enabled students in the match condition to acquire multiple perspectives on the problem and the problem-solving strategy, that are both seen as beneficial to problem-solving. Although more differences were expected, only one significant difference between the match and the causal condition was found.

Table 4: Multilevel Analyses concerning Students' Communicative Activities

| | Conceptual condition ($n = 10$) | Causal condition ($n = 10$) | Simulation condition ($n = 10$) | Match condition ($n = 6$) | Effects match condition ($N = 36$) | | |
|---|---|---|---|---|---|---|---|
| | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* | $\chi^2(3)$ | $\beta$ | *SE* |
| Coordination | 76.50 (37.58) | 60.62 (44.41) | 46.40 (17.33) | 149.40 (39.33) | 33.42 | 37.36* | 11.57 |
| Focusing | 16.92 (10.97) - | 12.38 (5.45) - | 8.20 (4.94) - | 33.40 (19.98) + | 24.11 | 8.48* | 3.06 |
| Checking | 40.42 (17.40) - | 32.88 (32.54) | 26.50 (9.08) - | 72.60 (7.83) + | 25.89 | 16.37* | 6.94 |
| Argumentation | 19.17 (13.12) | 15.38 (13.27) - | 11.70 (7.39) - | 43.40 (16.64) + | 26.21 | 12.19** | 3.49 |

Notes. * $p < .05$; ** $p < .01$; if match condition significantly > a mismatch condition than the match condition is indicated with a + and the mismatch condition with a -

## Problem-solving performance

One-way MANOVA on the total score of the problem-solving performance showed that groups in the match condition indeed scored significantly higher than groups in both the conceptual ($p = .02$; $d = 2.28$) and the simulation condition ($p = .05$; $d = 1.90$). Groups receiving a congruent ER for each part-task gave answers that (1) were more suited for a specific part-task, (2) contained more business-economics concepts and financial consequences, and (3) were more often correct. Although expected, no significant differences were found between the match and the causal condition.

# Discussion

Although based on 13 triads, the results of our study confirmed that providing part-task congruent guidance in the representational tools beneficially affects coordination processes and problem-solving performance. Groups receiving such support during their online collaborative problem-solving task were better able to coordinate their discussions about the concepts, principles and procedures of the knowledge domain than groups not receiving it. That is they more often (1) verified whether they are actually talking about the same discourse topic and, where necessary, repaired the common focus (i.e., focusing), (2) guarded the coherence and consistency of their shared understanding of the content space (i.e., checking), and (3) argued about their different points of view (i.e., argumentation). Those students were, therefore, better able to discus multiple perspectives on the problem and the problem-solving strategy (Hmelo-Silver, et al., 2007) and resulted in better answers to the part-tasks and better final solutions to the problem. Although the results seem very promising problem-solving performance of students in the causal condition was very similar to what was found in the match condition. Students in both conditions received the causal ER, that showed all relevant concepts, solutions and their causal interrelationships, providing the students multiple qualitative perspectives on the knowledge domain. It seems, therefore, important to support causal reasoning during collaborative problem-solving. Combining the causal ER with both other ERs might also hinder problem-solving when students experience difficulties integrating the different ERs. When students do not know how to combine multiple ERs, they might choose to stick with the familiar one and make no attempt to integrate the different ERs (Ainsworth, 2006). Additional research into the effects of representational scripting should be carried out to gain more insight into how students use and combine the concepts, solutions and relations within and between different ERs.

## References


Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16,* 183–198.

Andriessen, J., Baker, M., & Suthers, D. D. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M. Baker, & D. D. Suthers, (Eds.), *Cognitions: Arguing to learn* (pp. 1-25). Dordrecht, The Netherlands: Kluwer Academic Press.

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12,* 307–359.

Broeken, M., Jaspers, J., & Erkens, G. (2006). Virtual Collaborative Research Institute (VCRI). Version 2.3. Utrecht, The Netherlands: Utrecht University.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61-91). Heerlen, The Netherlands: Open Universiteit Nederland.

Erkens, G., & Janssen, J. (2008). Automatic coding of online collaboration protocols. *International Journal of Computer-Supported Collaborative Learning, 3,* 447–470.

Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior*, *21*, 463–486.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12,* 213–232.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42,* 99–107.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York/London: The Guilford Press.

Kirschner, P. A., Beers, P. J., Boshuizen, H. P. A., & Gijselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior*, *24*, 403–420.

Ploetzner, R., Fehse, E., Kneser, C., & Spada, H. (1999). Learning to relate qualitative and quantitative problem representations in a model-based setting for collaborative problem solving. *Journal of the Learning Sciences, 8,* 177–214.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning,1,* 315–337.

Suthers, D., Girardeau, L. and Hundhausen, C. (2003). Deictic Roles of External Representations in Face-to-face and Online Collaboration. Designing for Change in Networked Learning Environments. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds), *Proc. International Conference on Computer Support for Collaborative Learning* (pp. 173-182). Dordrecht: Kluwer Academic Publishers.

Van Bruggen, J. M., Boshuizen, H. P. A., & Kirschner, P. A. (2003). A cognitive framework for cooperative problem solving with argument visualization. In P. A. Kirschner, S. J. Buckingham-Shum, & C. S. Carr (Eds.), *Visualizing Argumentation: Software tools for collaborative and educational sense-making.* (pp. 25-47). London: Springer.

van der Linden, J. L., Erkens, G., Schmidt, H., & Renshaw, P. (2000). Collaborative learning. In P. R. J. Simons, J. L. Van der Linden, & T. Duffy (Eds.), *New learning* (pp. 1-19). Dordrecht: Kluwer Academic Publishers.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*, 1–30.

# Writing and commenting on professional procedures: In search of learning designs promoting articulation between school and workplace learning.

M. Gavota, M. Betrancourt, D. Schneider, University of Geneva, 1211 Genève 4, Switzerland
Email: Monica.Gavota@unige.ch, Mireille.Betrancourt@unige.ch, Daniel.Schneider@unige.ch

**Abstract:** The present study investigated the effect of scaffolding in a learning design in which apprentices were asked to reuse workplace experience to comment on peers' written productions. Apprentices were asked to perform an authentic workplace task, to explain how they got to the solution and then to comment on a peer's solution. Two conditions of scaffolding were compared: In the high scaffold condition, guiding questions were provided both for explanation and commenting, while in the low scaffold condition, a general instruction was given. The findings showed that, for the first tasks, the high scaffold group outperformed the low scaffold one with regard to the quality of the quotes and self explanations, but that the pattern reversed by the end of the activity. These results show that scaffolding is a useful tool for supporting reflection and learning, but that it should be adapted to the different steps of the activities.

## Theoretical background

According to Tynjala (2008), "the development of vocational and professional expertise requires the integration of different types of knowledge and interaction between theory and practice, and the development of the workplace as a learning environment both for employees and students is important to ensure the continuous development of competence. This requires close collaboration and partnership between education and work" (p.131).

In the case of vocational training, professional procedure acquisition represents a key issue in apprentices' professional development. Learning theories (e.g., Taatgen, Huss & Anderson 2008, Anderson, 1982) generally support the idea that direct teaching alone is not sufficient for the acquisition of procedural knowledge. Although largely untested, these theories suggest that teaching can only provide declarative knowledge, and that procedural knowledge should be acquired by the compilation of declarative knowledge through practice and feedback.

Nevertheless, learning theories focusing on procedure acquisition emphasize the role of direct instruction. Procedural knowledge development typically involves three steps (Weill-Fassina & Pastré, 2004). First, novices learn to apply a definite set of rules according to typical situations. In the second stage, workers learn to distinguish between classes of situations, particularly critical situations, and to behave accordingly. They form pragmatic "concepts", in the sense that they are oriented toward an improvement of the action performed, and not towards a better comprehension of the situation. Ochanine (1981) found that experts formed "operative images", which are simplified and biased representations of the objective situation, but are more relevant for the practice. For example, experts air traffic controllers typically under-estimate the distance between planes, which result in taking a larger security margin than strictly necessary (Bisseret, 1995). In the third stage, with the development of expertise, several operative concepts are formed and articulated, resulting in a more abstract representation of the task, which enables workers to handle efficiently a larger number of situations.

The main recent cognitive research in procedure acquisition still follows the central model ACT-R proposed and revised by Anderson (1993). In his revised model, ACT-R, Anderson (1993) focuses on the declarative memory of the examples demonstrating how the procedures should be performed. The first use of examples represents an analogy. Then, the production rules are compiled on the basis of this analogy. Moreover, it is assumed that the individual works on a declarative representation of the example. In other words, procedure acquisition would require a declarative representation of an example (but not necessarily in its long-term memory) because the procedure acquisition is done by learning from declarative traces of previous problem solving. This model emphasizes the importance of having a well developed declarative representation of the procedure in order to support the "compilation phase" meaning the passage from this declarative representation to the actual shortcut mastery of the procedure. In addition, according to Pirolli and Bielaczyc (1989), understanding the procedure's steps as well as the context to which the procedure is linked are essential elements for the procedure acquisition. It appears that the transferability of procedural knowledge is determined by a process of rules understanding (Kieras and Bovair, 1986). The initial acquisition of a procedure represents a process of understanding and not just a mimetic reproduction of the procedure's steps. The procedure acquisition is fostered by representing the procedure in its context, and understanding the different procedural steps in their own context (Broudy, 1977;

Bonner and Walker, 1994). Considering all those aspects of procedure acquisition, this study aims at finding means to support the construction of a strong declarative representation of a procedure as well as supporting the comprehension of its steps and the context in which it takes place.
Deeper processing of prior knowledge and examples represents a facilitating factor for supporting procedure comprehension and transfer (Chi, Bassok, Lewis, Reimann and Glaser, 1989). One effective way of avoiding superficial knowledge processing (King, 1992) and supporting deeper cognitive and metacognitive treatments - that students are likely to naturally avoid in the process of learning – is using external guiding under the form of prompts. These prompts are questions or hints supporting for efficient learning processes.
Metacognitive theories (e.g., Ge and Land, 2004) support the assumption hat becoming aware of oneself as a learner allows the student to reflect, monitor, and revise the process and products of his or her own learning. It was found that helping students develop abilities to monitor and revise their own strategies and to use resources may enable them to improve general learning expertise that can be used in a wide variety of settings (Kauffman, 2004). On the basis of this theoretical frame we designed and implemented learning tasks involving prompted self-explanations about the procedure, its steps and rationales.
Moreover, it has been shown (Hausmann and Chi, 2002; Chi, Bassok, Lewis, Reimann and Glaser, 1989) that both good students and prompted students tend to give more self explanations that refine and expand the conditions of an action, explain the consequences of an action, provide a goal for a set of actions, relate the consequences of one action to another and explain the meaning of actions.
Writing self-explanations about a procedure can thus be a suitable manner of supporting procedure comprehension as well as improving the organization of the declarative phase of procedure acquisition. In addition, the production of written text per se may be seen as a learning activity because it involves the production of a message based on domain knowledge and therefore domain knowledge should be "recovered", reorganized and incorporated into a linear form, a message understandable to someone else. Collaborative and peer writing/commenting activities represent an appropriate mean for reinforcing reflection and knowledge organization by the means of confrontation and hindsight taking in the reflection (Dillenbourg & Self 1994). Nevertheless, up to now, little research has investigated the effects of peer tutoring for "writing to learn" activities (Gielen, Dochy, Tops, Peeters, 2007). One critical issue is the level of scaffolding of such peer-commenting activity. As for self-explanation, it was assumed that peer commenting would benefit from guiding questions that orient towards argumentative and reflective writing, rather than on shallow and detail statements.

## The experiment

This paper reports a quasi-experimental study in which apprentices were asked to perform a task similar to a workplace situation (i.e. filling in a quote) and then describe how they got to the solution. Depending on the condition, self-explanation was either triggered by a general question (low scaffolding) or by guiding prompts (high scaffolding). They then were asked to comment on a peer's explanation, in the same scaffolding condition as their self explanation. Finally, they revised their quote taking into account the peer's comments.

We expected that prompted students would give more self-explanation statements (process measure) thus realizing deeper cognitive treatment than apprentices in low scaffolding condition, both for their own and for their peer's production. As a consequence, we expected that the "«high scaffold»" group would profit from deeper reflection activity and perform better in the two phases of the learning task (outcome measure).

## Methods

### Participants and context

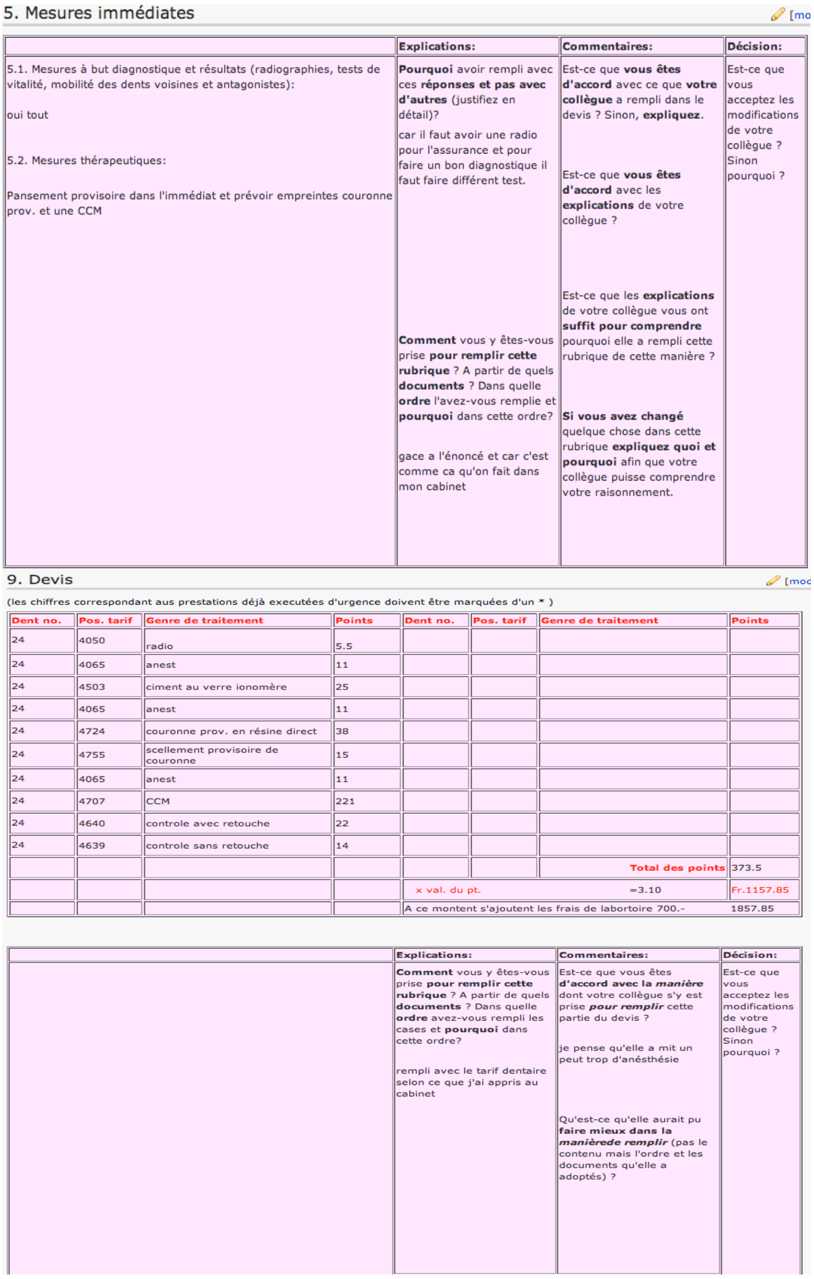
This activity took place in Geneva dental care vocational training school. It was designed in collaboration with the dentistry course teacher and was implemented during the school hours. The topic "filling in an insurance quote" was part of the curricula. Participants were 3rd year apprentices from two classes of the dental care assistant school in Geneva, all women[1]. Overall 26 students aged of 18-23 years participated in the experiment and were randomly allocated in one of the two experimental conditions (low or high scaffolding) according to a one by 2 between-subject design. The activity took two consecutive school hours.

### Material and apparatus

A wiki-based environment was used in order to support the apprentices' writing activities. An electronic equivalent of the quote form containing 4 tables (below, referred to as "table 3", "table 5", "table 7" and "table 9", see Figure 1) was proposed, each one referring to different aspects of the patient's dental problems, the medical treatment and the financial dimension. The "table 3" had to be filled in with information about the

[1] There were no men in this dental care assistant school.

damages caused by the accident (type of damage, teeth involved), the "table 5" referred to the diagnostic and treatment measures to be taken, the "table 7" enumerated the measures to be taken as a definitive treatment and finally "table 9" concerned the financial aspects (medical treatment, points, price and some other specific details).



5. Mesures immédiates [mo

| | Explications: | Commentaires: | Décision: |
|---|---|---|---|
| 5.1. Mesures à but diagnostique et résultats (radiographies, tests de vitalité, mobilité des dents voisines et antagonistes):<br><br>oui tout<br><br>5.2. Mesures thérapeutiques:<br><br>Pansement provisoire dans l'immédiat et prévoir empreintes couronne prov. et une CCM | **Pourquoi** avoir rempli avec ces **réponses et pas avec d'autres** (justifiez en détail)?<br>car il faut avoir une radio pour l'assurance et pour faire un bon diagnostique il faut faire différent test.<br><br>**Comment** vous y êtes-vous prise **pour remplir cette rubrique** ? A partir de quels **documents** ? Dans quelle **ordre** l'avez-vous remplie et **pourquoi** dans cette ordre?<br><br>gace a l'énoncé et car c'est comme ca qu'on fait dans mon cabinet | Est-ce que **vous êtes d'accord** avec ce que **votre collègue** a rempli dans le devis ? Sinon, **expliquez**.<br><br>Est-ce que **vous êtes d'accord** avec les **explications** de votre collègue ?<br><br>Est-ce que les **explications** de votre collègue vous ont **suffit pour comprendre** pourquoi elle a rempli cette rubrique de cette manière ?<br><br>**Si vous avez changé** quelque chose dans cette rubrique **expliquez quoi et pourquoi** afin que votre collègue puisse comprendre votre raisonnement. | Est-ce que vous acceptez les modifications de votre collègue ? Sinon pourquoi ? |

9. Devis [moc

(les chiffres correspondant aus prestations déjà executées d'urgence doivent être marquées d'un * )

| Dent no. | Pos. tarif | Genre de traitement | Points | Dent no. | Pos. tarif | Genre de traitement | Points |
|---|---|---|---|---|---|---|---|
| 24 | 4050 | radio | 5.5 | | | | |
| 24 | 4065 | anest | 11 | | | | |
| 24 | 4503 | ciment au verre ionomère | 25 | | | | |
| 24 | 4065 | anest | 11 | | | | |
| 24 | 4724 | couronne prov. en résine direct | 38 | | | | |
| 24 | 4755 | scellement provisoire de couronne | 15 | | | | |
| 24 | 4065 | anest | 11 | | | | |
| 24 | 4707 | CCM | 221 | | | | |
| 24 | 4640 | controle avec retouche | 22 | | | | |
| 24 | 4639 | controle sans retouche | 14 | | | | |
| | | | | | | Total des points | 373.5 |
| | | | | x val. du pt. | | =3.10 | Fr.1157.85 |
| | | | | A ce montent s'ajoutent les frais de labortoire 700.- | | | 1857.85 |

| | Explications: | Commentaires: | Décision: |
|---|---|---|---|
| | **Comment** vous y êtes-vous prise **pour remplir cette rubrique** ? A partir de quels **documents** ? Dans quelle **ordre** avez-vous rempli les cases et **pourquoi** dans cette ordre?<br><br>rempli avec le tarif dentaire selon ce que j'ai appris au cabinet | Est-ce que vous êtes **d'accord avec la *manière*** dont votre collègue s'y est prise ***pour remplir*** cette partie du devis ?<br><br>je pense qu'elle a mit un peut trop d'anésthésie<br><br>Qu'est-ce qu'elle aurait pu **faire mieux dans la *manièrede remplir*** (pas le contenu mais l'ordre et les documents qu'elle a adoptés) ? | Est-ce que vous acceptez les modifications de votre collègue ? Sinon pourquoi ? |

Figure 1. Screen capture of the wiki-based learning environment used in this study. From left to right: student A's answer, her self-explanation, student B's comment and A's final decision.

They were preceded by a yes/no question asking the apprentices about their previous familiarity at the workplace with this task.

Each table had two extra columns: one for students' explanations and the other one for the peers' comments. All apprentices were already familiarized with the use of the wiki tool since they already had used it for previous school activities. Depending on the condition, two versions of the material were designed, differing only on the content of the two extra columns of the quote: explanation and commenting. The "«high scaffold»" condition provided, for each of the tables, in the explanation column, a set of five questions guiding their answers. The questions were built on the 5W criteria (why, when, who, where, how) and used the Lin and Lehman's (1999) "reason justification", "rule based" types of prompts. In the «low scaffold» condition, there was only the general prompt: "Explain".

## Procedure

The apprentices were welcomed by the teacher, as for a normal class. In the first phase, starting from an authentic case study proposed by the teacher, the apprentices had to fill in an electronic quote and explain how they did it. In the high scaffolding condition, the apprentices had to explain why they had filled in that way, how they did to fill in, on which material/knowledge they based themselves and finally in which order they filled in the table and why. The «low scaffold» group had the same task but for the explanation column they didn't have any scaffolding questions, they only had the instruction to "Explain". In a second phase, a colleague had the task to correct a peer's quote and also comment about the changes she had done and explain, using the same level of scaffolding as for the first phase. The "high scaffold" group had a list of specific questions aiming to see if they agreed with the way the table had been filled in, if they had changed/corrected anything and why and if they considered the explanations given by the peer satisfying, while the "«low scaffold»«low scaffold»" group had only the instruction to "Comment".

### Data analysis

For each of the quote's tables we analyzed the percentage of filling in the quote, the explanations and the comments, as well as the quality for each of these three writing tasks.

For the percentage of filling in, we coded with 0 when there was nothing written, 1 when the table was filled in at 30%, 2 when it was at 50%, 3 at 75% and finally 4 representing a completely filled in table, by taking into account the number of fields to be filled in. These codes concerned only the fact of having filled in the table regardless of the entries' correctness.

When coding the quality of the quotes, we counted, in each table, the number of correct answers, the number of acceptable answers (representing neither correct nor incorrect answers) and the number of incorrect answers. We also calculated the sum of correct/acceptable/wrong answers for the entire quote, in its first and second version. With regard to the quality of the explanations, we counted the number of valid arguments, the number of details mentioned, as well as the number of incorrect explanations for each one of the four tables. Comments we classified into two categories, informative and non-informative. The comments classified as informative were those providing useful information to the author for improving the quote while the non-informative ones were the social ones (containing support, acknowledgement etc.)

## Results

In the first place, we checked out if there was any influence of the apprentices' declared familiarity with the task at the workplace and their performances at the quote activity in the classroom. We conducted an ANOVA, which showed no significant differences between the quality of the "familiarized" apprentices' quotes and the quotes of the ones declaring not being familiarized with this procedure ($F(1,23) = 4.85$, $MSE = 21.33$; $p < .05$).

Following our hypotheses, we looked for the differences between the two experimental groups with respect to 1) the proportion of filling in each of the quote's tables; 2) the quality of the entire quote as well as table by table; 3) the quality of the explanations; 4) the quality of the comments.

### Percentage of information entered in the quote

We looked at the number of information completed in the entire quote in its first and second version, as well as each of the four tables separately, regardless of their accuracy.

The proportion of completion of the entire quote was computed as the ratio of the number of information entered against the total number of answers required for each apprentice. An ANOVA analysis did not reveal any significant differences between the two experimental groups ($p > .05$), neither in the first, nor in the second versions.

Considering each table separately in the first and second version, the ANOVA analysis revealed one significant difference – the high scaffold group filled in more in the first version of Table 3 than the low scaffold group ($F(1,24) = 8$; $MSE = 22.15$; $p < .05$). Nevertheless, this difference faded away in the second version of the quote where no significant difference between the groups was found anymore.

### Correctness of the quotes

The quality of the quotes was evaluated through the number of correct answers to the first and second version of the quote for each apprentice. ANOVA test revealed that the low scaffold group produced better quality quotes than the high scaffold group in the first version ($F(1,24) = 4.82$; $MSE = 36.9$: $p < .05$) as well as in the second version of the quote ($F(1,19) = 5.82$; $MSE = 35.9$: $p < .05$).

A deeper analysis concerning the differences between the two groups with regard to the correct and wrong answers in each of the quote's four tables revealed interesting information.

In the first draft, the high scaffold group produced significantly better answers for the table 3 than the low scaffold group ($F(1,24) = 6.54$; $MSE = 0$; $p < .05$). Howver, we did not find any significant difference between

the groups in the second version of the table 3 – meaning that the «low scaffold» group compensated the difference after the revision of the quote.
Significant differences were also found in the quality of table 9. In the first as well as in the second version of this table, the low scaffold group gave significantly more correct answers than the high scaffold group (first version: $F(1,24) = 7.14$; $MSE = 30.15$; $p < .05$ second version: $F(1,19) = 4.58$; $MSE = 24.31$; $p < .05$). Moreover, the high scaffold group gave more incorrect answers in both versions of the table 9 (first version: $F(1,24) = 11.53$; $MSE = 3.84$; $p < .01$, second version: $F(1,19) = 7.13$; $MSE = 4.86$; $p < 0.5$). These results mean that participants in the high scaffold group did not take advantage of the comments on their first version to improve enough in the second draft in the extent that they would equalize or outperform the answers of the low scaffold group.

### Differences between the first and second versions of the quotes

To evaluate the evolution of the quotes from the first to the second version, the difference between the number of correct answers over the two versions of the quote was computed for each condition. Neither the high nor the low scaffold group significantly improved the quality of their quotes from their first to their second versions. Nevertheless, if we look at the observed results, both the low scaffold and the high scaffold group improved the quality of the quotes between the first and the second version (Low scaffold: $M_{first} = 9.08$, $M_{second} = 9.82$; High scaffold, $M_{first} = 6.69$, $M_{second} = 7.2$).

### Quality and quantity of the self-explanations

In terms of percentage of completing the explanations, ANOVA analysis did not reveal any significant difference between the two groups, meaning that the two groups produced more or less the same amount of explanations.
Nevertheless, there were significant differences between the two groups in the quality of the explanations. The high scaffold group gave significantly more arguments for the table 3 than the low scaffold group ($F(1,24) = 6$; $MSE = 6.5$; $p < .05$), but it also gave significantly more "details" ($F(1,24) = 7.38$; $MSE = 2.46$; $p < 0.5$). For the table 9, the situation was the exact opposite: the low scaffold group gave marginally significantly more arguments than the high scaffold group ($F(1,24) = 3.64$; $MSE = 26$; $p = .06$) but significantly less details ($F(1,24) = 4.52$; $MSE = 1.88$; $p < .05$).
Regression analysis performed on the entire data, without regard to the experimental condition, showed that more argumentative explanations given in table 3 ($Rsquare = .23$; $F(1,25) = 7.3$; $p < 0.5$) and table 9 ($Rsquare = .31$; $F(1,25) = 10.76$; $p < .01$), was related to better accuracy of these tables.

### Type of Comments

For each group, comments were simply classified into two categories, according whether they were superficial or essential. The number of comments in each category was used to evaluate the quality of the comments. On the entire quote, we found a marginal significant difference in the favor of the low scaffold group, who made more essential comments than the high scaffold group ($F(1,20) = 4.26$; $MSE = 10.8$; $p = .053$). Analyzing table by table, we notice that the low scaffold group systematically gave more essential and less superficial comments than the high scaffold group, but none of these differences was statistically significant. We can only consider them as tendencies.
Regression analyses did not reveal any statistical significant relations between the quality of the explanations and the quality of the comments or between the quality of the quote and the quality of the comments.

## Discussion

Apprentices' declared familiarity with the procedure did not seem to account for their actual performance, tested with the class activity. This might be explained by the apprentices' low self-evaluation competencies but also by the very heterogeneous workplace training they get. These conjectures reinforce the need to implement school activities based on apprentice's professional experience as well as supporting knowledge sharing and collaborative reflection.
With regard to the effect of the scaffolding, the results partly contradicted our hypothesis by showing that for some tasks, the group with low scaffolding outperformed the group with high scaffolding. Having a synthetic look at the results, we noticed that the high scaffold group outperformed the low scaffold group at all levels (quality of the table, percentage of filling in the table, quality of explanations) in the starting task (table 3). The two groups have similar performances on the middle tasks (table 5 and 7, not reported here). When analyzing the final task (table 9), the low scaffold group has significantly better results (at all levels) than the high scaffold group.
Considering the quality of the self-explanations proposed by the two groups, we notice that even though they seem to produce the same amount of information, the high scaffold group produced more valid arguments in the explanations in the first table but ended up by proposing more superficial ones in the final table. The low

scaffold group had exactly the opposite behavior. Statistical analysis also confirmed that the quality of the explanations was a predictor for the quality of the quote.

These results support the idea that high scaffolding may have helped in the beginning of the activity in order to support good explanations and reflection (King, 1992), generates more correct answers, and orients the apprentices towards appropriated task solving approaches. Nevertheless, by the end of the activity, the low scaffold group caught up with the high scaffold one and even outperformed it. One could argue that high scaffolds can be useful in the beginning of the task, in order to orient the apprentices' cognitive strategies and give indications about how the task should be dealt with, but, shortly after becomes superfluous or even too burdening. This study underlines the idea that scaffolding is a powerful tool for fostering deep cognitive treatment of knowledge in self-explanation task on procedures. Nevertheless, it identifies the need of conceiving and implementing adaptive designs in which these prompts fade out along the task in order to give place at the adaptive and learning skills of the student to take the place.

However, it should be noted that the low scaffold group started with better performance in terms of global accuracy of the first version of the quote. No data in our study can disentangle the effect of scaffolded self-explanation and initial level of each group.

Regarding the effect of peer commenting, the results could not confirm the positive effect of this instructional methods through statistical analysis, though the observed results showed an improvement in performance in both scaffolding condition. It may be that self-explanation in itself has such a great instructional effect that the effect of peer-commenting could not be easily demonstrated using few measures with few students.

Last but not least, we have to take into account that this study took place in real class settings, and that other factors may have interfered. Students were not very familiar with self-explaining activities nor commenting ones. They might not have fully benefited by the instructional device (e.g., Ainsworth & Burcham, 2007). It is nevertheless a real situation, valuable for conceiving and implementing these types of designs in real vocational training classes. Future long term studies will investigate further this issue as well as the relation between the computer support and the commenting instruction.

## References

Ainsworth, S., & Burcham, S. (2007). Impact of text coherence on learning by self-explanation. *Learning and instruction, 17*(3), 286-303

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Erlbaum.

Bisseret, A. (1995). *Représentation et décision experte - Psychologie cognitive de la décision chez les aiguilleurs du ciel*. Toulouse : Octares.

Bonner, S., Walker, P. (1994). The Effects of Instruction and Experience on the Acquisition of Auditing Knowledge, *The Accounting Review, 69*(1), 157-179.

Broudy, H. (1977). Types of knowledge and purposes of education. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 1–17). Hillsdale, NJ: Lawrence Erlbaum.

Chi, M.T.H., Bassok M., Lewis M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Dillenbourg, P., & Self, J. (1992). People Power: A human-computer collaborative learning system. In C. Frasson, G. Gauthier, and G. McCalla (Eds.) *The 2nd International Conference on Intelligent Tutoring Systems*, *Lecture Notes in Computer Science* (pp 651-660). Berlin: Springer-Verlag.

Ge, X., & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development, 52*(2), 5-22.

Gielen, S., Dochy. F., Tops, lL., Peeters, E., (2007). Effects of formative peer assessment on writing performance: What is the most beneficial role of the assessee. in *Proceedings of the 12th EARLI biennal conference on Research on Learning and Instruction,* Aug. 28 - Sept. 1, Budapest (Hungary).

Hausmann, R., & Chi, M., (2002). Can a computer interface support self-explaining? *The International Journal of Cognitive Technology, 7*(1), 4-14.

Kauffman, D. F. (2004). Self-regulated learning in Web-based environments. Instructional tools designed to facilitate cognitive strategy use, metacognitive processing and motivational beliefs. *Journal of Educational Computing Research, 30*(1&2), 139-161.

Kieras, D. E., & Bovair, S. (1986). The acquisition of procedures from text: A production-system analysis of transfer of training. *Journal of Memory and Language, 25*, 507–524.

King, A. (1992). Facilitating Elaborative Learning through Guided Student-Generated Questioning, *Educational Psychologist*, 27 (1), 111-126.

Lin, X.D., & Lehman, J. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching, 36*(7), 1–22.

Ochanine, D. (1981). L'image opérative, *Actes du séminaire et recueil d'articles*. Paris : Université Paris 1.

Pirolli, P. and Bielaczyc, K. (1989). Empirical Analyses of Self-Explanation and Transfer in Learning to Program. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 450–457). Ann Arbor: MI.

Taatgen, N.A., Huss, D. & Anderson, J.R. (2008). *The Acquisition of Robust and Flexible Cognitive Skills. Journal of Experimental Psychology: General, 137*(3), 548-565.

Tynjala, P., (2008). Perspectives into learning at the workplace, *Educational Research Review, 3*, 130–154.

Weill-Fassina, A., & Pastré, P. (2004). Les compétences professionnelles et leur développement. In P. Falzon (Ed.), *Ergonomie*. Paris: PUF.

## Acknowledgments

This study was conducted within the *Dualt leading house* (Technologies for dual vocational education) funded by the Swiss Federal Office for Professional Education and Technology.

# Complexity, Robustness, and Trade-Offs in Evaluating Large Scale STEM Education Programs

Susan A. Yoon, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104
yoonsa@gse.upenn.edu
Lei Liu, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104
leil@gse.upenn.edu

**Abstract:** This study explores the application of the complex systems lenses of robustness and trade-offs to evaluate the success of a STEM education program in terms of sustainability. A mixed methods approach is used to document contextual constraints, trade-offs considered such as modularity vs. one-size-fits-all, and actions taken by the project team to build toward adaptive capacities.

## Introduction

The issue of declining participation of America's youth in science, technology, engineering and math (STEM) education and careers has received a great deal of attention over the last few years from industry, government, and education sectors (Business Roundtable, 2005; Domestic Policy Council, 2006; U.S. Department of Education, 2007). As STEM jobs continue to grow, by 2012 the number of positions in science and engineering is estimated to outpace the number of qualified people to fill them by 26% and 15% respectively (NSF, 2006). From a citizenship science perspective, there is a broader issue of a general lack of STEM literacy in society associated with such declining participation. When fewer students are interested in and aware of emerging STEM research such as nanotechnology, the likelihood that they are able to contribute to intelligent decision making about applications that can affect their daily lives decreases (Yoon, 2008). Additional pressures on workforce development exist with the emphasis on developing 21st century and cyberinfrastructure-enabled scientific skills in which learning from and using digital technologies both for developing conceptual knowledge and process skills must now figure prominently in education (NSF, 2007; Partnership for 21st Century Skills, 2007). To respond to these needs researchers in the Learning Sciences and STEM education have focused attention on understanding how to develop, sustain and scale innovative science and technology reform-oriented programs. Investigation and discussion have centered on what mechanisms and variables are required to create the conditions for success and how success may be evaluated (Dede, Honan & Peters, 2005; Yoon & Klopfer, 2006; Yoon et al., 2009). One theme that resonates through much of this research is the challenge reformers face in adapting their programs to specific contexts (Datnow, 2005; Dede & Honan, 2005; Hargreaves & Fink, 2000). Elmore (1996) writes about the difficulties experienced by nested clusters of innovation in educational settings. He states that failures, historically, in generating successful large-scale reforms can be attributed to an "absence of practical theory that takes account of the institutional complexities that operate on changes in practice" (p. 21). Similarly, Goldman (2005) and Fishman et al. (2004) identify the need to account for multiple embedded levels of stakeholders when designing for improvement, who must actively support the reform if it is to succeed. Coburn (2003) reinforces the idea that educational reform and improvement are matters of complexity highlighting the inability of research to address the inherent multidimensionality between and within educational constituents. Building from this reform literature and the calls for research methods and practical theories that take a systems approach, in this paper we introduce the principle of robustness, both as a lens and a tool for understanding issues of sustainability and scale of innovative STEM education programs. We suggest that achieving a robust programmatic state is arguably the sine qua non of success with respect to sustainability, which in turn provides the rationale for going to scale. As we will argue, one of the core measurable mechanisms operating in the service of robustness is trade-offs. We present data, analyses, and results of the first year of a three-year STEM program in which we document the trade-offs experienced and then undertaken in the redesign process to create a more robust program.

## Theoretical Framework

### Complex Systems

Multidimensionality, nested clusters, and multiple constituents and levels are all characteristics of complex systems. The study of complexity has been a focus of research in many academic disciplines such as biology (Kaufmann, 1995), physical science (Bak, 1996; Prigogine & Stengers, 1984), psychology (Arrow et al., 2000), anthropology (Lansing, 2006), and economics (Stacey, 1996). The ubiquity of complex systems research appears to extend from the commonalities that exist between structures or contexts and behaviors in that they comprise multiple elements that adapt or react to patterns that they create (Arthur, 1999). By studying the patterns that emerge and the interactional dynamics that lead to these patterns, researchers can better understand among other things, how systems adapt, self-organize, fluctuate and either achieve or do not achieve steady states. For example, in the simple dilemma of the Tragedy of the Commons, one individual may choose to act selfishly and use up resources with no real impact on supplies. However, if many individuals decide to behave in the same way soon the resources will be depleted and the individuals along with the collective community will suffer. Thus the culture or the collective that exists must be considered. It requires one to know the limits of the culture and to adjust behaviors in order to sustain a robust system, which is also in the best interest of individual survival.

### Robustness in Complex Systems

Related to understanding the limits of the culture, robustness researchers study the contextual parameters that enable systems to persist or deteriorate in the face of negative often, unpredictable impacts such as the spread of viruses in an ecosystem, continual military conflict between countries, or the collapse of stock markets in economies. As a central complex systems mechanism, various problems across biological and social situations can be investigated through the lens of robustness which has led leading complex systems organizations like the Santa Fe Institute to devote whole research tracks to its study (http://www.santafe.edu/research/topics-innovation-evolutionary-systems.php#4). One of the unifying characteristics that enable its multidisciplinary application is that perturbations impact natural, cultural, and engineered systems. It is hypothesized that studying varying system's responses to these perturbations can provide general rules or principles that can be used to solve long-standing questions such as how evolution that requires variation supports the emergence of phenotypes from genotypes (Wagner, 2005). Others have used notions of robustness to understand how the structure of social networks enables information flow and resource management (Webb & Bodin, 2008). For example, as individual nodes, actors or elements are removed from the network key connections may be broken which may isolate subgroups and thus limit how information gets distributed. Elsewhere we have used social network theories, analyses, and tools to help students and teachers develop better information-seeking strategies to construct knowledge and to gain social capital (Yoon, accepted; Baker-Doyle & Yoon, 2009). In all these studies, creating robust systems requires ease of information flow in order to give and receive resources. Where constraints exist due to cultural or contextual barriers, trade-offs are needed to build adaptive capacities.

### Trade-Offs

Perhaps the most well known example of a general application of the idea of trade-offs is the situation of the Prisoner's Dilemma that falls under the general topic of Game Theory. This is a hypothetical case where two prisoners are interrogated separately for having together committed a crime. In the classic scenario, it is assumed that the base prison sentence is 5 years. The pay-offs or decrease of prison time will depend on whether the individuals choose to cooperate with each other or defect. In the context of the game, defecting (claiming innocence and implicating your partner), while your partner cooperates (confesses to the crime), will always provide the largest pay-off to the defector in that no prison time will be served. Similar to the Tragedy of the Commons example, using this simple algorithm, researchers have studied how it is possible that cooperation might have emerged as an evolutionarily adaptive strategy (Axelrod, 1984). While a full blown discussion of the mechanisms that lead to cooperation cannot be made here, one of the key ideas related to trade-offs for our research pertains to the fact that much of the study of strategies that allow populations rather than individuals within systems to benefit are underpinned by the fact that the success of someone or something that includes activities and resources may come at the expense or sacrifice of another person or another thing. For example, Janssen and Anderies (2007) write with respect to the trade-offs that organizations must make in choosing to satisfy short-term or long-term goals, "when all ‘low-hanging fruit’ is taken to increase robustness cheaply, [systems] will eventually reach a point at which it is no longer possible to generate additional robustness without a cost to performance and/or decreased robustness somewhere else in the system" (p. 44).

### Complexity, Robustness, and Trade-Offs in Evaluating STEM Education Programs

Understanding the educational system as a complex system is not new in educational research. In his *Change Forces* series focused on reform, Fullan (1993, 1999, 2003) uses complex systems theory as an organizing framework to reveal core concepts such as non-linearity, unpredictability and multi-level agency that are important issues to contend with in real-world educational systems. However, where states of existence like robustness become goals that drive the reform efforts rather than for description, and where mechanisms like trade-offs become actions, educational research takes a step closer to a "practical theory" that accounts for institutional complexities that Elmore (1996) searches for. The following study documents the multiple constituents, contextual limits, and important trade-offs made toward increasing robustness in a large-scale STEM education reform effort.

## Methodology

### Context

This work is funded under the National Science Foundation program "Innovative Technology Experiences for Students and Teachers (ITEST)". The ITEST program is designed to increase opportunities for students and teachers in underserved schools to learn and apply information technology concepts and skills in the STEM content areas. Our project is a school district-university partnership between one of the largest school districts in the U.S., a school of education and a leading university nanoscale research center. It aims to achieve the broader ITEST goals by updating standard high school science curricula and training teachers through a curriculum and instruction framework built on five component variables addressing content knowledge, pedagogical content knowledge, and workforce development goals. These variables are: 1) Real world science and engineering applications such as current nanotechnology research and ethical considerations of this research; 2) Educational Technologies to develop content knowledge such as NetLogo computer simulations; 3) Information Technologies for communication, community-building and dissemination such as Google Groups and wikis; 4) Cognitively-rich pedagogical approaches such as problem-based learning; and 5) Investigation and consideration of STEM education and careers.

There are two parts to the scope and sequence of project activities. The first part entails a three-week, 75-hour professional development summer workshop for science teachers, in which they learn to construct and pilot curricular units based on the five component variables. These curricular units are intended to be aligned with school district standards for high school biology and physical science. The summer workshop is followed by the school-year implementation of these units in teachers' classrooms and 5 follow-up Saturday professional development workshops. In addition to the central curriculum and instruction goals, another major goal of the project is to build teaching capacities and self-directed interests through communities of practice structures which include working with groups of teachers from the same school, implementing an online comprehensive professional development database for peer-to-peer interaction, and requiring teachers from different schools to construct units collaboratively. For this study, we focus on the activities and program developments of the first cohort of teachers who participated on the project between August 2008 and May 2009.

## Participants

Ten male and six female teachers participated in the workshop from 10 high schools and 1 middle school in the district. The group was racially/ethnically diverse: seven teachers were White, six were African American, and three were Asian. Courses taught ranged from grades 8 to 12 in the content areas of physical science, biology, chemistry, and physics. The average amount of teaching experience was 15.8 years, with a range of 1 to 39 years of experience. Data from 128 students taught by a sample of six teachers were collected with the following racial/ethnic breakdown: 23.4% White, 48.4% African American, 7.8% Hispanic, 14.1% Asian, and 3.9% other races. The study also investigated perceptions and actions of the project team, as the goal was to document and understand the kinds of trade-offs we experienced in the development and implementation of the program. Thus, we include the project team as participants who comprised: a professor, a post-doc and two doctoral students from the school of education; and a professor, two post-docs, and the educational outreach director from the nanoscale research center.

## Data Sources and Analyses

Eight data sources were collected for the study and analyzed through a mixed methods approach.

1. A 64-item 5-point Likert-scale survey administered to teachers to measure teacher's self-perceptions of pedagogical practices, beliefs, student participation, and teacher confidence levels with particular emphasis on the project's five variable curriculum and instruction framework. A pre-intervention survey was collected in August 2008 and a post-intervention survey was collected in May 2009. Complete data sets were obtained from 8 participating teachers due to various data collection challenges. A paired-samples t-test was conducted to determine if teachers had statistically significant change in their self-perceptions from the beginning to the end of the 2008-2009 academic year.
2. An 82-item 5-point Likert-scale survey administered to students to measure self-perceptions of attitudes toward science and use of classroom resources and strategies with particular emphasis on the project's five variable curriculum and instruction framework. Pre- and post-intervention surveys were collected from 5 participating teacher's classrooms immediately before and after the ITEST implementation. Times for survey collection were different due to the variability of implementation in the school-year curriculum. For example, some teachers chose to develop units in physical science on the topic of properties of matter, which corresponded with delivery of the standard curriculum at the beginning of the year. Others chose units that corresponded with latter times. A paired-samples t-test on 128 survey responses was conducted to determine if students had statistically significant change in their self-perceptions.
3. An 11-item open-ended research survey administered to teachers that probed in-depth understanding of nanoscale content and STEM education pedagogical beliefs related to the 5 variable project framework. A pre-intervention survey was collected in August 2008 and a post-intervention survey was collected in May 2009. Complete data sets were obtained from 8 participating teachers.

4. Qualitative case studies of 5 project teachers were conducted by the project's external evaluator. Case study classrooms were selected to obtain information from as diverse a sample as possible. The case-study implementation was guided by the following 4 questions: What does the ITEST curriculum look like as it is implemented in case-study classrooms?; To what extent do teachers implement the key elements of ITEST?; To what extent is each enacted unit similar or different from the intended unit?; What supports and challenges are observed and reported in the implementation of the ITEST curricular units? Data collection included: field notes from three classroom observations, informal conversations with teachers, informal conversations and/or focus groups with students, and a formal end-of-unit interview with each teacher.
5. Focus group interview for the entire cohort of teachers conducted during the Saturday workshop held at the end of February 2009. Focus group questions probed for details about affordances and constraints to implementation and suggestions for program redesign for the next cohort of teachers and schools. The cohort was divided into 3 smaller groups. Interviews were transcribed and compared by the external evaluator and the project team in order to identify common themes.
6. Weekly 1.5 hour project management meetings audio-recorded and transcribed.
7. Written and informal verbal reflections of the 4 scientists from the nanoscale science center after specific nano-content meetings and after content modules were implemented during the summer and Saturday professional development workshops.
8. Field notes and informal verbal reflections of the 4 school of education researchers documented throughout the year of implementation as facilitators in participant teachers' classrooms. There were 33 formal classroom observation field notes collected.

For data sources 3, 5-8, data were reviewed systematically by the 4 school of education researchers through a qualitative grounded theory approach (Strauss & Corbin, 1998) in which all open-ended surveys, transcribed meetings, observation notes, and reflections were mined for evidence illustrating various trade-offs that emerged throughout the project year.

## Results

With respect to increasing robustness and revealing and using trade-offs that are specific to our ITEST project, we were interested in documenting the emerging themes as they pertained to the 5 variable framework as well as impacts on the goal of building a practitioner community focused on the framework. In this paper, the results of two variables (i.e., real world science and engineering applications and educational technology) are presented and organized into the three sub-categories of: i) Contextual Constraints; ii) Trade-offs; and iii) Actions. We continue to work toward identifying more general categories of constraints and trade-offs that may be applied across similar STEM education programs, however, in this section we list them individually and attempt to apply a meta-level categorization later in the Discussion section.

### Real world science and engineering applications

Research in nanotechnology refers to the interdisciplinary study of physical, chemical, and biological phenomena in the 1 – 100 nanometer range. Manipulation of atoms and molecules at this scale have given scientists unprecedented abilities to engineer materials, products and processes in, medicine, the environment, cosmetics, clothing, and computer and automotive technologies (Roco, 2003). Nanotechnology research has undergone rapid growth over the past decade with an estimated annual global research expenditure of $13.9 billion (PCAST, 2008). The pervasiveness of these applications is also immense. The Woodrow Wilson *Project on Emerging Nanotechnologies* states that there are over 800 products currently on the market produced by over 400 companies in 21 countries. Therefore, in our project, incorporating current and cutting edge nanotechnology research and applications including nanoscale content and real world problem solving into the standard science curriculum was a central design element of the teacher constructed PBL units.

#### Contextual Constraints

Several contextual constraints limited the success of this variable. First from the case studies and focus group interviews, the most widely reported challenge teachers faced was the mismatch in PBL pedagogy and the school district's system of managed instruction, i.e., core curriculum and benchmarks. Due to time challenges, delivering a complete unit was not possible for some teachers choosing instead to deliver discrete lessons throughout the year. Despite significant increases in teacher confidence levels in their ability to use problem-based instructional practices ($t(7)=2.546$, $p=.038$), teacher effectiveness in helping students identify multiple solutions to a problem ($t(7)=2.121$, $p=.078$), and perceptions of increased use of real world applications in the curriculum ($t(7)=3.667$, $p=.008$), no significant differences were found in student perceptions of learning about real world issues, new science innovations or discoveries, teacher's demonstrations of science problems, and student problem-solving. Moreover, while teaching aspects of PBL units was found to occur in classroom observations, even in the best example of one of the case-studies, student final projects had low fidelity with the original curriculum, which included plans for students to gather, synthesize, and evaluate the various types of information used throughout the unit in order to assess the ethical issues related to their nano-application.

Difficulties also existed in aligning curricular content levels and school contexts with nanoscale applications. Despite significant increases in teacher's perceptions of their ability to incorporate nanoscale science content ($t(7)= 3.130$, $p=.017$), scientists frequently expressed concerns about finding the appropriate connections to the topics covered in the high school curriculum. Issues for the scientists also surfaced with respect to empowering teachers to "own" the presented concepts so that they felt compelled to shape it into instruction relevant for their classrooms. From a content knowledge perspective, responses from teachers that probed understanding of nanoscale concepts showed misconceptions at the fundamental level of size and scale as well as in their perceptions of nanoscale science requiring an entirely new understanding of canonical scientific domains. For example, the majority of teachers believed that nanotechnology is a distinctively new science (with its own set of laws), instead of an addition to existing sciences, like biology, chemistry, and physics.

## Trade-Offs

The trade-offs we would like to highlight here deal with structural and behavioral factors of social and human systems, which we have discussed elsewhere to be among the most difficult constraints to contend with in the educational system (Yoon & Klopfer, 2006; Yoon et al., 2009). In the first case regarding the implementation of PBL, the expectation was for teachers to develop a coherent unit with the higher-order inquiry skills of evaluation and synthesis however, given district testing pressures, the actualized curriculum looked much different. This is akin to the trade-off of robustness to performance. Essentially, the trade-off parameter can be described in terms of <u>some benefits are better than none</u> in the face of highly uncompromising intransigent constraints. In the second case in aligning curricular content and the existence of fundamental misconceptions, if we were to follow the research on how people learn or knowledge-building theory (Bransford et al. 2000; Scardamalia, 2002), providing scaffolds for teachers to construct their own knowledge would stand a better chance in helping them make curricular connections on their own, promote self-efficacy and self-regulation, and remediate misconceptions. However, the major constraints here are lack of time and resources to undertake the knowledge-building process. A logical trade-off parameter then, given that teachers felt positive about their ability to incorporate nanoscale content into their regular curricula, might be to <u>meet the teachers half-way</u>, i.e., providing instructional support and curricular resources to seed interest and a solid leverage point to explore research and application avenues other than the ones that are directly taught to them.

## Actions

Based on the trade-off parameters, a number of changes in project expectations and program delivery occurred for cohort 2 teacher professional development. Instead of mandating a coherent sequential unit, curricula could be constructed drawing on content connections in multiple topics throughout the school year. We identified a teacher from cohort 1 who developed particular expertise in working with project goals in this way and we enlisted her to teach a two-hour module on what we called semi-coherent curricular units where she modeled the process and provided to teachers her own tried and true unit with detailed explanation of the important modifications she made in practice. Actions spawned by the trade-off of meeting the teachers half-way, included constructing a comprehensive curriculum alignment inventory in which 11 nanotechnology and concepts/applications were identified in addition to how they interface with the State Standards and the School District's Core Curriculum for grade 9 Physical Science and grade 10 Biology. Table 1 depicts a sample of the inventory the project team developed. This was used in instruction for teachers when new content was introduced and given to them as a general heuristic to use when building their standards-aligned PBL units.

<u>Table 1: Sample of Curriculum Alignment Inventory.</u>

| Nano/Bio Concepts/Applications | Pennsylvania State Standards | | SDP Core Curriculum |
|---|---|---|---|
| | Physical Science | Biology | |
| 1. What is the nanoscale? How small is nano?<br><br>2. Unique properties at the nanoscale | 3.4A Explain concepts about the structure and properties of matter<br><br>3.1D. Apply scale as a way of relating concepts and ideas to one another by some measure. | 3.3A Explain the structural and functional similarities among living things | Physical Science (PS) 3 – Atoms & Periodic Table, & Chemical Bonds<br>BIO 2- Introduction to Chemistry |
| 3. Nanomaterials and surface area | 3.4B. Analyze energy sources and transfers of heat.<br>• Determine the efficiency of chemical systems by applying mathematical formulas.<br>• Use knowledge of chemical reactions to generate an electrical current. | 3.3B. Describe and explain the chemical and structural basis of living organisms.<br>• Describe relationships between the structure of organic molecules and the function they serve in living organisms.<br>• Identify the specialized structures and regions of the cell and the functions of each. | PS 2 – Matter<br>BIO 3 – Cell Structure and Function<br>BIO 2- Introduction to Chemistry |

## Educational Technologies to develop content knowledge

Educational technologies (ET) were presented to teachers as distinct in our framework from information technologies (IT) in that ET comprises imaging, simulations, and visualization tools that help scientists and students learn about science content and processes whereas IT is primarily used for communication and interaction. Teachers were instructed on the emerging importance of being able to navigate the cyberinfrastructure for STEM careers, which includes using computational tools for data collection, manipulation, visualization, and prediction. Several digital tools used with teachers included simulations found on the Internet such as The Nano Journey, http://www.nanoreisen.de/english/index.html and NetLogo (Wilensky & Reisman, 2008) software in which a drug delivery system for nanotechnology and cancer was constructed and used with teachers to learn about how to conduct scientific experiments with such simulations.

### Contextual Constraints

In comparison to using information technologies, this variable was much less prevalent in teacher's instructional practices. Classroom observation field notes indicated that only 2 out of the total cohort of 16 teachers used NetLogo in their instruction. Where as teachers surveys showed significant increases in the perception of their ability to incorporate IT into their lessons ($t(7)=2.198$, $p=.064$), no differences were found with ET. In student surveys, although the item investigating science or other ideas through simulations, images, or animations ($t(125)=2.594$, $p=.011$) was found to have positive gains, actual observed practice showed no use of these tools for higher order scientific inquiry skills. Rather, they were used mainly to illustrate or demonstrate a concept. In the case studies, similar results were found along with other contextual constraints. One major challenge was that the teachers did not have administrative privileges to download NetLogo onto their school computers. One teacher did download NetLogo onto his laptop and students individually came up to his computer and tried the simulation for short periods of time. Another teacher said that she was interested in NetLogo, but didn't have a laptop that she could load it onto. Of the other three, one said that he didn't have enough facility to use it, and the other two did not think it would have been feasible to use in their classes due to classroom management issues. All teachers did use other educational technology, the most common application being visualization software that allowed students to move between visualizations of the same item or location at different scales.

### Trade-offs

The lack of educational technology use triggered several in-depth discussions during project management meetings about how to facilitate greater applications as well as applications that resembled more closely how scientists really used such software for scientific discovery. NetLogo and similarly StarLogo were simulation tools used by the principal investigator in previous projects with some success in developing computational literacies and 21st century skills (Klopfer & Yoon, 2005; Klopfer et al., 2005). Particularly with respect to the potential for manipulating variables, changing initial conditions, and running multiple experiments, NetLogo/StarLogo modeling software could be extremely useful tools in the science classroom. However, given the difficulties in downloading the software, differences in teachers facility and comfort levels, and the fact that some ET was being used in instruction, the project team explored the trade-off of modularity vs. one size does not fit all. This trade-off is similar to the use and benefits of differentiated instruction. Although it was believed that more focused practice on a few ET applications for all teacher participants would work toward successful ET implementation, again, structural and behavioral realities posed seemingly insurmountable constraints and we needed to differentiate this component.

### Actions

Several actions were taken according to the identified trade-off. First, with the easily accessible Internet visualization programs, we developed exemplar lessons to show generically how experimentation could be done with any similar tool. Next, we devoted a good deal of time collecting and vetting ET sites on the Internet that would be appropriate for our project which culminated in a comprehensive Educational Technology Inventory with clickable links. We also planned to deliver parallel sessions that demonstrated different ET applications and associated pedagogies for which teachers could choose to sign up according to their interests and comfort levels. We would offer these iteratively so that when teachers felt that they had developed particular competencies, they could choose to attend increasingly more challenging ET modules to improve their skills.

We have documented and used several more trade-offs as a response to contextual constraints. For example, after realizing how little access we had to guidance counselors in schools and coming to the understanding that they were no longer assigned to just one school, plans to work on a comprehensive nanotechnology education and career package in which teachers and guidance counselors would collaborate to disseminate on a large scale needed to change. Instead we opted for different STEM education and career experiences, one of which was a short, real-world internship where teams of two high school students worked in a modestly paid after school job for two hours a week in the nanoscale professor's research lab. The trade-off and extensive project team discussions centered around project impact on individual vs. many students. With the variable of cognitively-rich pedagogies, a trade-off that we continually face is how much is enough theory vs. practice. In working toward our goal of building a community of practice one of the major trade-offs we grapple with is whether, how, and how much to build technological, human and/or social capital.

## Discussion

With all of these trade-offs, one common theme that organizes around building sustainable or robust programs appears to be negotiating the middle ground. When working with systems with high complexity, that is, many often competing factors that operate under unpredictable conditions, finding the compromise between ideal project fidelity and real-world constraints may be the key to ensuring that at least some parts of the vision survives. From this platform, we can create opportunities for continued growth of activities that have taken a hold in the education system to continue to grow. Another meta-level theme that clearly impacts the project team's decision-making practices is the disposition of flexibility. The adaptive process that these STEM education programs necessarily undergo can work if there is an expectation of flexibility and a willingness to carefully document and analyze the constraints in order to choose another path that may be more successful. Likewise, adopting an ideological stance of pragmatism may be a prudent choice. If we apply from the start, the expectation that ideas or theories such as how people learn best can only be measured or deemed successful in practice, we are forced to take account of the institutional complexities that can exert heavy influence on program success.

There are at least three potential contributions that this study can make. At the level of theory understanding sustainability of innovative STEM education programs through the lenses of complexity, robustness and trade-offs is a promising framework to evaluate the complexity of educational reform. The study also informs practice, operationalizing the framework in terms of documenting contextual constraints, trade-offs considered, and then actions taken. Finally, looking across the trade-offs and identifying the common themes of finding the middle ground, developing dispositions of flexibility, and adopting a pragmatic stance, can be used by other like-minded educational reformers in the design of future STEM education programs.

## References


Arrow, H., McGrath, J. & Berhahl, J. (2000). *Small Groups as Complex Systems*. Thousand Oaks, CA: Sage Publications, Inc.

Arthur, W.B. (1999). Complexity and the economy. *Science* (284), 107–109.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality*. New York: Copernicus.

Baker-Doyle, K. & Yoon, S. (2009, August). In search of practitioner-based social capital: A social network analysis tool for understanding and facilitating teacher collaboration in a professional development program. Paper to be presented at the annual meeting of the *European Association for Research on Learning and Instruction*, Amsterdam, Netherlands.

Bransford, J., Brown, A., & Cocking, R. (Eds.) (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Business Roundtable (2005). *Tapping America's Potential: The Education for Innovation Initiative.* [Web Resource] Available online. www.businessroundtable.org/publications.

Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3-12.

Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Educational Administration Quarterly*, *41*(1), 121-153.

Dede, C., Honan, J. & Peters, L. (2005). Scaling Up Success: Lessons Learned from Technology-Based Educational Improvement. San Francisco, CA: Jossey-Bass.

Dede, C. & Honan, J. (2005). Scaling up success: A synthesis of themes and insights. In, C. Dede, J. Honan & L. Peters (Eds.) *Scaling Up Success: Lessons Learned from Technology-Based Educational Improvement* (pp. 227-239). San Francisco, CA: Jossey-Bass.

Domestic Policy Council (2006). *American Competitiveness Initiative*. Office of Science and Technology Policy: Washington D.C.

Fishman, B., Marx, R., Blumenfeld, P., Krajcik, J. & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *Journal of the Learning Sciences*, *13*(1): 43-76.

Goldman, S. (2005). Designing for scalable educational improvement. In, C. Dede, J. Honan & L. Peters (Eds.) *Scaling Up Success: Lessons Learned from Technology-Based Educational Improvement* (pp. 67-96). San Francisco, CA: Jossey-Bass.

Hargreaves, A. & Fink, D. (2000). Three dimensions of educational reform. *Educational Leadership*, *57*(7), 30-34.

Janssen, M & Anderies, J. (2007). Robustness trade-offs in social-ecological systems. *International Journal of the Commons*, *1*(1), 43-65.
Johnson, S. (2001). Emergence: The Connected Lives of Ants, Brains, Cities, and Software. New York: Touchstone.
Kauffman, S. (1995). *At Home in the Universe*. New York: Oxford University Press.
Klopfer, E. & Yoon, S. (2005). Developing games and simulations for today and tomorrow's tech savvy youth. *Tech Trends, 49*(3), 33-41.
Klopfer, E., Yoon, S. & Um, T. (2005). Teaching complex dynamic systems to young students with StarLogo. *Journal of Computers in Mathematics and Science Teaching*, 24(2), 157-178.
Lansing, S. & Miller J. (2005). Cooperation Games and Ecological Feedback: Some Insights from Bali. *Current Anthropology 46*(2): 328-334.
National Science Foundation. (2007). Cyberinfrastructure Vision for 21st Century Discovery. Arlington, VA: Author.
President's Council of Advisors on Science and Technology (PCAST). (2008). *The National Nanotechnology Initiative: Second Assessment and Recommendations of the National Nanotechnology Advisory Panel*. Washington, D.C.: Executive Office of the President.
Prigogine, I. & Stengers, I. (1984). *Order Out of Chaos*. New York: Bantam Books.
Roco, M. (2003). Converging science and technology at the nanoscale: Opportunities for education and training. *Nature Biotechnology*, *21*(10), 1247-1249.
Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal Education in a Knowledge Society* (pp. 67-98). Chicago: Open Court.
Strauss, A. L. & Corbin, J. M. (1998). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (2nd edition), Sage, Newbury Park, CA.
U.S. Department of Education (2007). *Report of the Academic Competitiveness Council*. Washington, DC: Author.
Wagner, A. (2005). *Robustness and Evolvability in Living Systems*. Princeton, NJ: Princeton University Press.
Webb, C. & Bodin, O. (2008). A network perspective on modularity and control of flow in robust systems (pp. 85-118) . In, J. Norberg and G. Cumming (Eds.) *Complexity Theory for a Sustainable Future.* New York: Columbia University Press.
Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and Instruction, 24*(2), 171–209.
Yoon, S. (accepted). Using social network graphs as visualization tools to influence peer selection decision-making strategies to access information about socioscientific issues. *Journal of the Learning Sciences*.
Yoon, S. (2008). An evolutionary approach to harnessing complex systems thinking in the science and technology classroom. *International Journal of Science Education*. 30(1), 1-32.
Yoon, S., Chessler, M., Kumar, R., Williams, D., Flicker, J. & Dunham, S. (2009, April). Understanding emergent challenges and adaptations in the design of an OST STEM project through a Structure, Behavior and Function (SBF) Complexity Lens. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Orange County, CA.
Yoon, S. Liu, L. & Goh, S. (2009). Exploring the process of convergent adaptation in technology-based science curriculum construction. In the proceedings of the *Computer Supported Collaborative Learning annual conference*, Rhodes, Greece.
Yoon, S. & Klopfer, E. (2006). Feedback (F) Fueling Adaptation (A) Network Growth (N) and Self-Organization (S): A Complex Systems Design and Evaluation Approach to Professional Development. *Journal of Science Education and Technology*, *15*(5-6), 353-366.

## Acknowledgments
This research is supported by a National Science Foundation Grant: ESI 0737437.

# Exploring Convergence of Science Ideas through Collaborative Concept Mapping

Dana Gnesdilow, Anushree Bopardikar, Sarah A. Sullivan, & Sadhana Puntambekar
University of Wisconsin-Madison
gnesdilow@wisc.edu, bopardikar@wisc.edu, sasullivan2@wisc.edu, puntambekar@education.wisc.edu

**Abstract:** This exploratory study examined how groups of sixth grade students worked together to create a collaborative concept map. We selected two contrasting cases based on their initial maps, one heterogeneous group with divergent maps and one homogeneous group with convergent maps. We analyzed group dialogue and collaborative and individual concept maps to understand: 1) if convergence of science ideas occurs during collaborative concept mapping and 2) how convergence or divergence during collaboration influences individual map construction. We found that collaborative concept maps facilitated greater convergence of ideas in groups with initially divergent pre individual maps. Further, the negotiation of divergent science ideas during collaboration led to gains in individual students' science understanding. Implications of findings and directions for future research are discussed.

Concept maps have been used in a variety of educational settings to support meaningful learning (e.g., Daley, et al. 2008) in contexts such as reading (e.g., Nesbit & Adesope, 2006), writing (Sharples, 1994), and measuring conceptual change (e.g., Edmondson, 2000). Concept maps represent knowledge by focusing on relationships between ideas. They consist of concepts delimited by circles and arrows linking two concepts (Novak & Cañas, 2008). Words placed on the arrow explain relationships between concepts.

The use of concept maps has been mainly informed by the cognitivist premise that learning occurs within the individual through the assimilation of new concepts and relationships within existing propositional frameworks. A concept map represents an individual's knowledge structure at a given point in time (Fisher, 2000). However, Roth and Roychoudhury (1994) point out that "the theoretical framework in which concept mapping is grounded is not much concerned with the social construction of knowledge" (p.438). They describe concept mapping as a conscriptional device for social thinking through engaging in discourse, and an inscriptional representation of the groups' shared understanding of concepts and their relations. Further, they assert that new knowledge is constructed in collaborative peer groups, and refer to prior research to claim that this knowledge is taken up by the individual participants through engagement in this social interaction.

Several other researchers have explored collaborative concept mapping activities from a socioconstructivist perspective, and have reported that collaborative concept mapping has potential to support student learning (see Basque & Lavoie, 2006, for a review). In particular, van Boxtel, van der Linden, Roelofs, and Erkens (2002) found in an experimental study that students involved in collaborative concept mapping engaged in elaborate conceptual discourse and co-constructed meaning. They also found that elaborate student discourse influenced individual learning outcomes. Although collaborative mapping research suggests learning benefits, Teasley and Fischer (2008) contend that collaborative activities do not always facilitate individual learning or ensure equivalent learning gains for all individuals. They further state that few studies have explored how collaborative convergence of ideas relates to individual learning outcomes. This issue of convergence and divergence in thinking is critical in collaborative learning. For instance, Stahl (2004) stresses the importance of initially divergent ideas during collaboration. Further, Schwartz (1999) argues that individuals create unique understandings as they attempt to co-construct shared meaning during collaboration.

There is limited research that has investigated what a group map as an inscription represents; that is, whether there is convergence (the construction of shared understanding) in ideas between students as they collaboratively construct a concept map in a classroom, and the effect of collaborative mapping on individual understanding of concepts. In this exploratory study, we examined the level of convergence during collaborative concept mapping. From a socioconstructivist perspective, we also explored if a concept map represents the ideas of an entire group and how collaborative concept mapping relates to individual student understanding. Our premise is that a movement from shared understanding to construction is seen when students generate their own understanding in a new individual context based on previous interactions with others. To document this premise, we employed individual and group concept maps as inscriptional products, and student audio during collaborative concept mapping as a conscriptional process. We first examined the group and individual concept mapping scores of five groups from one science class. We then selected two contrasting cases based on their initial maps, one heterogeneous group with divergent maps and one homogeneous group with similar maps, to explore how the two groups worked together to create a collaborative concept map. Our research questions were: 1) Is there convergence toward a shared understanding of science concepts during a collaborative concept map activity? 2) How does convergence or divergence during the collaborative concept map influence individual map construction?

# Methods

## Participants and Instructional Context

The study was conducted in a sixth grade science class in a private Midwestern school. Students engaged in a design-based science curriculum to learn about simple machines using the CoMPASS hypertext system to complete a set of design challenges (Puntambekar, Stylianou, & Goldstein, 2007). In this six week curriculum, students brainstormed predictions and questions, conducted research on the CoMPASS hypertext system, and conducted hands on investigations to test their ideas in the same small groups throughout the unit. Students completed five mini design challenges pertaining to simple machines. Students also engaged in several individual and collaborative concept map activities throughout the unit. Practice individual and collaborative maps were constructed during the first half of the unit followed by the pre map in the middle of the unit. Students then constructed post individual maps at the end of the unit. Collaborative concept maps were created at the end of each mini design challenge.

For this study, we used data from the pre and post individual concept maps and students' collaborative maps. Students worked for approximately 45 minutes to construct paper and pencil individual concept maps involving simple machines physics concepts. Students were instructed to draw a concept map with at least nine (on pre map) or 14 (on post map) physics concepts with a description of how the concepts were related. A focus on how the machines work was emphasized in these instructions. In contrast, the collaborative concept map served both as a conscriptional device to aid students in their physics learning and as a group product or inscription of their learning for assessment. For the collaborative concept map, constructed between the pre and post individual concept maps, students were instructed to create a paper and pencil concept map incorporating physics ideas related to pulleys. They worked face-to-face on these maps for approximately 15-20 minutes during their regular science class.

## Data Sources and Analysis

We calculated concept map depth ratios to capture the sophistication of science ideas of students in all five groups. Further, we examined the map layout and coded group dialogue captured while students generated their collaborative maps for two contrasting groups, selected on the basis of their initial individual maps.

### Individual and Group Concept Map Depth Ratios

For this part of the analysis, we focused on sophistication of concept map propositions. A proposition is two concepts connected by a linking word to form a semantic unit. Modified from prior work (Puntambekar, Stylianou, & Goldstein, 2007), we scored all concept maps based on 1) the number of accurately described concept propositions and 2) the sophistication of science ideas expressed in the propositions. We assigned proposition scores according to a five point scale. This scale ranged from: -1 to 3 (see Table 1). We then calculated a *depth ratio* for each concept map. The depth ratio was calculated by dividing the sum of the scores for each of the propositions on a map by the total number of propositions on the map. A higher depth ratio value signifies more sophisticated understanding of the relationships between concepts. We chose to use the depth ratio to measure student progress because it considers the depth of the propositions regardless of the number of propositions on the map, and gives an average score of the sophistication of all propositions. Two researchers coded 15% of all individual and group concept maps and achieved approximately 90% interrater reliability. One researcher scored the remaining maps.

Table 1: Concept Maps Scoring Rubric.

| Score | Description | Example |
|---|---|---|
| -1 | Incorrect | levers are inclined planes |
| 0 | Ambiguous language (has, gives, uses, needs) in reference to concepts (MA, force, friction, distance) | screw has effort force |
| 1 | Fact, type of, is a, example, overgeneralization | third class is one of three kinds of lever |
| 2 | Definition, affects, intuitive language, increases, decreases | friction reduces MA |
| 3 | Scientific language, elaborate explanations, specify conditions for increase or decrease | levers increase the MA when the fulcrum is closer to the load |

### Individual and Group Concept Map Layout

We compared the map layout of both individual and group maps. We examined the chosen root word and structural organization (relational vs. hierarchical) between individual pre and post maps, and group maps. The chosen root word, or most central concept on a map, could be a machine or a physics concept. A relational concept map differs from a hierarchical map because relationships could be vertical as well as lateral, whereas in a hierarchical map relationships are primarily vertical.

### Group Dialogue

We transcribed the audio of student dialogue that occurred during the collaborative concept mapping activity. Total audio data consisted of approximately 35 minutes of audio and seven pages of transcripts. We inductively developed a set of nine codes after a preliminary examination of transcripts to capture the convergent and divergent exchanges between group members as they talked about science to construct their group map (see Table 2). Our codes also align with the transactive knowledge convergence process described by Weinberger, Stegmann, and Fischer (2007). We coded the transcripts at the utterance level. Each utterance could be assigned multiple codes.

We examined percentages of group dialogue to understand both the patterns of overall interactions in each group and the contributions offered by individual students. We calculated percentages by dividing the number of utterances categorized into a particular code by the total number of utterances coded during the collaborative activity. Similarly, the percentages for types of individual talk were calculated by dividing the number of utterances coded into a category by the total number of coded utterances of a particular student. Two authors independently coded 70% of the transcripts and established an interrater reliability of approximately 83%. The first author subsequently coded the remaining transcripts.

Table 2: Coding Rubric for Examining Collaborative Concept Mapping Audio Data.

| Code: | Description |
|---|---|
| Initiation of Ideas (II) | bringing up a new science idea |
| Simple Sharing (SS) | stating a science idea not taken up by other group members |
| Agreement of Ideas (AI) | explicit agreement between at least two group members |
| Contention of Ideas (CI) | explicit disagreement between group members in the form of questions or statements |
| Resolution of Disagreement (RD) | explicit consensus after a disagreement |
| Raising Questions (RQ) | asking group members questions about science concepts |
| Extending Ideas (EI) | refining or elaborating upon other group member's ideas |
| Restating Ideas (RI) | student restates idea or summarizes many group exchanges |
| Negotiation of Map Construction (NMC) | dialogue pertaining to what ideas to include on the map and how to position and connect them |

## Results

We report our findings based on three types of data: 1) individual and group map depth ratios, 2) individual and group map layout, and 3) individual contributions and group interactions, to understand if convergence of ideas occurred and how convergence was related to students' performance on the post individual concept maps.

### Depth Ratio Comparisons between Individual and Group Concept Maps

In this section, we discuss concept map depth ratio scores on students' pre and post individual maps and group map scores. We also discuss learning gains made by students as seen in the difference between pre and post map depth ratio scores. We calculated learning gains by dividing the actual gain by the total possible gain from pre to post maps. As we discuss the depth ratio scores, it is important to understand what small differences in the ratios represent. The depth ratio gives an average score of all the propositions on a map. A depth ratio of 1.1 can be best understood as consisting of mostly propositions falling into a level one score (see Table 1). Since the depth ratio has a small range from negative one to three, small changes in scores are meaningful in terms of student learning gains and differences in scores.

In Group A, the pre individual map depth ratios suggested convergence in their, with a difference of only .09 between the highest and lowest scores in the group. Their collaborative map, with a depth ratio of 1.0, reflected a level of understanding similar to what students showed on their individual pre maps. We also found that Rose's depth ratio decreased on her post individual map and she exhibited no learning gain. Both Alex and Lincoln performed slightly better on their post individual maps, with learning gains of 21% and 13% respectively. Group A's average learning gain was 8.6%.

The difference in pre individual map scores in Group B was most divergent at 0.85. In contrast with Group A, their collaborative concept map showed a modest improvement in their depth of understanding (1.39) over their individual pre map scores. Unlike Group A, all students in Group B made substantial gains in their depth ratio scores on the post individual maps with learning gains ranging from 36% to 51%. The average learning gain for Group B at 41% was much higher than any other of the four groups.

The second most divergent group was Group C with a difference of .73 between students' pre individual depth ratio scores. Like Group B, Group C had a higher collaborative map depth ratio score at 1.4 than the other groups. Group C also had the second highest average learning gains at 16% out of all five groups, with all students in the group showing learning gains. Although their average learning gain is smaller than that of Group B's, it is twice the gain than that of Groups A, D, and E.

Overall, Group D and E had depth ratio differences and learning gains similar to those of Group A, including lower depth ratio scores on their collaborative concept maps (at .93 for Group D and .64 for Group E). Like Group A, Groups D and E were relatively more convergent on their pre individual map scores with a difference in scores of .32 and .28 respectively. The average learning gains of Group D at 6.6% and of Group E at 5.8% were slightly lower than that of Group A at 8.6%. Like Rose in Group A, Jesse in Group D exhibited no learning gain and Brian in Group E made a small learning gain.

Thus out of the five groups, Group A had the most convergent initial maps, and Group B had the most divergent map levels before the collaborative concept mapping activity. We therefore examined student dialogue in these two contrasting groups to better understand how students created a collaborative concept map, and how their collaboration affected their individual learning.

Table 3: Depth Ratio Scores and Percent Learning Gains of Students' Pre and Post Individual and Group Maps.

| Group | Student | Pre Individual Map | Difference on Pre | Group Map | Post Individual Map | Percent Learning Gain or Loss: Individual Maps | Average Group Percent Learning Gain: Individual Maps |
|---|---|---|---|---|---|---|---|
| A | Alex | 1.00 | .09 | 1.0 | 1.41 | 21% | 8.6% |
| | Rose | 1.09 | | | .94 | -8% | |
| | Lincoln | 1.06 | | | 1.35 | 13% | |
| B | Betty | 0.44 | .85 | 1.39 | 1.29 | 36% | 41% |
| | Keesha | 1.08 | | | 1.77 | 37% | |
| | Jake | 1.29 | | | 1.60 | 51% | |
| C | Kristen | 1.46 | .73 | 1.4 | 1.60 | 7% | 16% |
| | Jerry | 1.10 | | | 1.33 | 13% | |
| | Mary | 1.08 | | | 1.39 | 15% | |
| | Collin | .73 | | | 1.17 | 20% | |
| D | Naomi | 1.12 | .32 | .93 | 1.30 | 10.5% | 6.6% |
| | Jesse | 1.29 | | | 1.16 | -5.8% | |
| | Amy | 1.00 | | | 1.24 | 10% | |
| | Armando | 1.32 | | | 1.50 | 11.8% | |
| E | Allison | 1.22 | .28 | .64 | 1.32 | 5.6% | 5.8% |
| | Brian | 1.05 | | | 1.10 | 0.03% | |
| | Gage | 1.33 | | | 1.48 | 11.8% | |

## Map Layout Comparisons between Individual and Group Maps for Groups A and B

Examining contrasting cases shows potential for understanding how collaborative processes influence learning outcomes (Rummel & Hmelo-Silver, 2008). Alex, Rose and Lincoln were the students in the more convergent group, Group A. While Betty, Keesha, and Jake were the students in the more divergent group, Group B. In Group A, Alex drew the collaborative concept map and in Group B, Betty drew the group map.

Groups A and B present interesting contrasts in their group and individual map layouts (see Table 4). We first examined the root word choices on pre individual, collaborative, and post individual concept maps. We found that all students used a machine as a root word on their pre individual map. Group A chose a machine root word, whereas Group B used a physics concept. Rose and Lincoln in Group A continued to use a machine as a root word on their post individual map, but Alex chose a physics concept. In contrast, all students in Group B changed from using a machine root word to using a physics concept root word on their post map, consistent with what had taken place in the collaborative mapping activity.

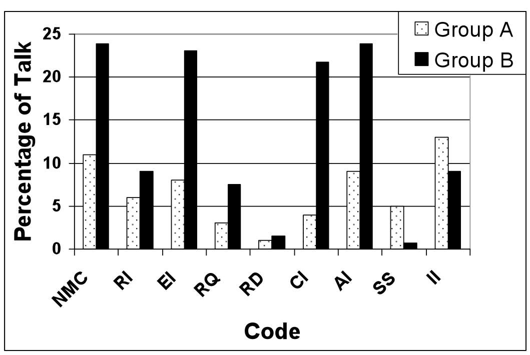
Second, we examined the structural organization of individual and group maps. Our analysis revealed that both groups drew relational group maps. Rose and Lincoln from Group A, and all students from Group B, had constructed hierarchical individual pre maps. Rose and Lincoln continued to draw a hierarchical map, but all students in Group B used a relational structure on the post individual map.

Table 4: Map Layout of Individual Students' in Group A and B's Pre and Post Individual and Group Maps.

| | | Root Word of Map: Machine (M) vs. Physics Concept (PC) | | | Map Structure: Hierarchical (H) vs. Relational (R) | | |
|---|---|---|---|---|---|---|---|
| Group | Student | Pre | Group Map | Post | Pre | Group Map | Post |
| A | Alex | M | | PC | R | | R |
| A | Rose | M | M | M | H | R | H |
| A | Lincoln | M | | M | H | | H |
| B | Betty | M | | PC | H | | R |
| B | Keesha | M | PC | PC | H | R | R |
| B | Jake | M | | PC | H | | R |

## Patterns of Group Dialogue and Individual Contributions in Groups A and B

We analyzed group dialogue to further understand the convergence or divergence that may have occurred when students engaged in collaborative concept mapping. This analysis was done to a) examine if students' ideas converged through engaging in collaborative interactions, and b) to see how convergence or divergence during the collaborative concept map activity related to individual map construction. These findings will enable us to understand how aspects of group dialogue may possibly explain the changes between students' pre and post individual concept maps. First, we will report on general discourse patterns for each group and compare the two groups (see Figure 1). Subsequently, we will present a more detailed analysis of each individual's major contribution to the group dialogue (see Figure 2).




| Code Legend | |
|---|---|
| NMC | Negotiation of Map construction |
| RI | Restating Ideas |
| EI | Extending Ideas |
| RQ | Raising Questions |
| RD | Resolution of Disagreements |
| CI | Contention of Ideas |
| AI | Agreement of Ideas |
| SS | Simple Sharing |
| II | Initiating Ideas |

Figure 1. Collaborative Concept Map Dialogue in Groups A and B.

In Group A, the most prominent pattern was initiating ideas (12.5%), agreement of ideas (8.75%), and extending ideas (7.5%). The percentage of talk related to the negotiation of map construction, was 11.25%. In looking at individual contributions to the collaboration, Alex's main contributions were agreement of ideas (20.69%) and extending ideas (13.79%). Rose's major contributions were initiating ideas (25%) and negotiating map construction and restating ideas (16.7%). Lincoln made a limited contribution to concept related talk (12.5%). Overall, the percentage of science related talk in this group during the collaborative concept mapping activity was relatively low, because a good portion of Group A's utterances were off task or not applicable to the study.

In contrast, Group B engaged in relatively more science talk during the collaborative concept mapping activity than Group A. Their dialogue mainly involved agreement of ideas (23.9%), extending ideas (23.1%), and the contention of ideas (21.7%). The percentage of talk related to negotiation of map construction was 23.9%. When examining individual contributions to the discourse, we find that Betty's talk focused on the contention of ideas (25.6%) and extending ideas (25.6%). Similarly, Keesha engaged in the group discourse through agreement of ideas (32.5%) and contending ideas (20%). Jake's principal involvement in the group process was mainly extending ideas (31%), but he also emphasized the negotiation of map construction (29.1%).

We found clear differences between the two groups' patterns of talk. Group A focused less on science, whereas Group B was focused on negotiating science ideas and negotiating map construction. Furthermore, the individual contributions from students in the two groups were also different. In Group A, students engaged less with the collaborative concept mapping and engaged in more talk unrelated to the activity and science. Students in Group B talked more about science concepts and contributed in diverse ways to their group map. Finally, a large proportion of talk in Group B was spent in contending ideas, unlike in Group A.

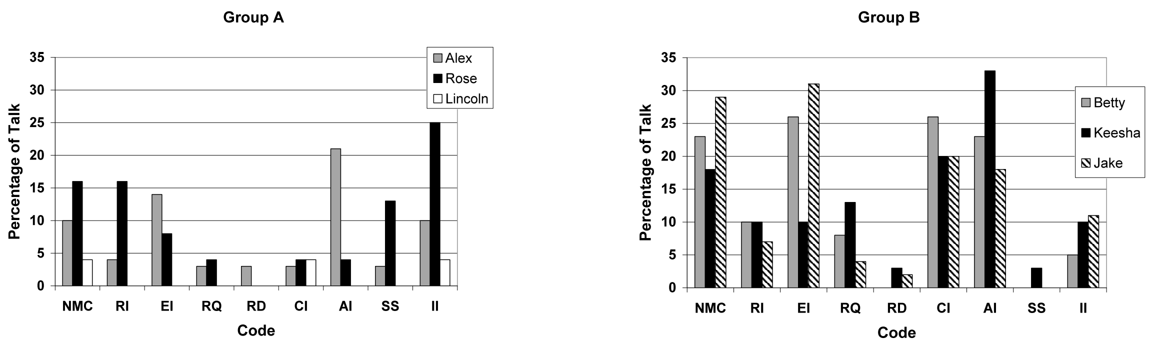



Figure 2. Individual Contribution to Collaborative Dialogue in Groups A and B.

## Discussion

Our exploratory study contributes to the literature on concept mapping and collaborative learning because it examines group maps both as conscriptional *processes* and inscriptional *products*. Our study is consistent with Daley et al's. (2008) call for research to understand how groups construct knowledge, and how this knowledge construction aids in group performance. It also aligns with Basque and Lavoie's (2006) emphasis for further investigation of the relationship between collaborative learning and concept mapping, particularly the types of interactions that promote learning.

In investigating whether a collaborative concept map serves as a conscriptional device in the classroom, we found differences between both the product and the process of collaborative concept mapping in the two contrasting groups. Our results indicate that there was less convergence between students in Group A than in Group B during the collaborative process. Roth and Roychoudhury (1994), as well as van Boxtel, van der Linden, Roelofs, and Erkens (2002), claim that collaborative concept mapping promotes shared understanding among students. However, our findings show that whereas a collaborative concept map may act as a conscriptional device to facilitate convergence of ideas, not all groups benefit equally from this activity when used in the classroom.

Students in Group A had pre individual maps that initially showed convergence. The negotiation and contention of science ideas and map construction between these students during collaborative concept mapping was relatively lower than Group B, resulting in a less sophisticated group concept map product, as evidenced by a lower depth ratio score and simple machines root word choice. Derbentseva, Safayeni, and Cañas (as cited in Novak & Cañas, 2008) found that root word choice affects the quality of a concept map. A physics concept root word may promote the creation of propositions focused on *relationships,* resulting in more sophisticated maps. Having a simple machine as a root word may lead to shallower, fact-based links. Further, Alex drew the map and was the only student in the group who made a relational map on the pre and post individual map. Despite having made a relational group map, Rose and Lincoln created hierarchical individual post maps. Ruiz-Primo and Shavelson (1996) have emphasized that relational maps provide a richer representation than a hierarchy, indicating more sophisticated maps. Group A's group map may reflect Alex's thinking more than that of the whole group, which could explain the low convergence in map layouts on the individual post maps.

The dialogue in Group A provides further evidence of a low level of convergence. We found that students in Group A did not participate equally in this process. Alex and Rose engaged in some negotiation of the science concepts, but there was little convergence on science ideas between group members. For example, of the ideas that Rose initiated many were also categorized as simple sharing, because they were not extended by the group, indicating limited mutual engagement as they discussed the science. Furthermore, Lincoln made little contribution to the science discourse. Thus, our findings suggest low convergence of science ideas in Group A.

Conversely, there was more convergence of science ideas among students in Group B. They created a more sophisticated group map than Group A, as measured by a higher depth ratio score, a physics concept root word and a relational organization. Students converged on their post individual map layouts, because all of them changed from using a machine-based root word to using a physics concept root word, and from a hierarchical to a relational structure, consistent with their group map.

Group B's dialogue provides further evidence for convergence. Group B had about twice as much dialogue about negotiation of map construction as Group A. These students contributed in diverse ways to the dialogue, which was more varied and involved more contention and negotiation of science concepts than Group A. Perhaps this dialogue contributed to both the changes and convergence of the map layout on individual post maps.

Our findings suggest that the level of convergence achieved during collaborative concept mapping could have influenced students' performance on the post individual maps. The students in Group A engaged less in constructing a shared understanding of the science ideas. Unlike van Boxtel et al. (2002), we found that students who actively contributed to the discourse did not always exhibit high learning gains. The post map depth ratios showed that Alex and Lincoln had greater learning gains than Rose's, whose depth ratio decreased. While Rose did engage in the science discourse, many of her contributions were not taken up, perhaps resulting in her less sophisticated post individual map and negative learning gain. Thus, the divergence in the level of engagement of ideas during the collaborative mapping activity in Group A could have played a role in their learning, resulting in differences in individual map layouts and learning gains on their post maps.

In Group B, the higher level of convergence during collaboration might have contributed to post map learning gains for all students and to more sophisticated map layouts by all three students in the group. For example, all of the students in Group B changed their root word to a physics concept and changed to creating a relational post individual map after having worked on a relational group map, suggesting greater convergence and sophistication. The higher level of convergence in Group B could explain the substantial learning gains of all individual students in the group. We contend that the remarkable growth of these students collectively and individually can be understood in light of Stahl (2004) and Schwartz's (1999) argument. First, Stahl (2004) suggests that divergent ideas between group members have a significant impact on collaboration. Because Group B started with more divergent individual maps, each student brought divergent ideas to the collaboration. The students also negotiated their diverse perspectives to construct a shared understanding. Individuals construct novel understandings as they attempt to create shared meaning during collaboration (Schwartz, 1999). Students in Group B constructed a sophisticated individual understanding after their collaboration, as seen in their substantial individual learning gains. This notion is further substantiated by examining data from the other three groups. Like Group A, Groups D and E, who were initially more convergent, made the least learning gains out of the five groups and created less sophisticated group maps. Alternately, like Group B, Group C, the second most initially divergent group, had twice as much average learning gain and made a more sophisticated group map than the three more convergent groups. These trends lend further credence to the notion that initial divergence between group members is important for productive collaborations and positive individual outcomes.

These results have important implications for the classroom, because the role that collaborative concept mapping plays as a conscriptional and inscriptional activity becomes complex in a classroom setting. For example, in this study, students did not contribute equally to the group activity, unlike in the experimental studies of van Boxtel et al. (2002). Also, students made unequal gains in their post depth ratio scores, a finding that is consistent with Teasley and Fischer's (2008) argument that collaborative activities do not ensure equal learning gains for all students. One area of concern is that the overall depth ratios are relatively low on individual post maps for students in all five groups. This indicates that there is room for improvement in the depth of science talk and learning gains. Further, our results question the validity of using collaborative concept maps to assess group understanding, because they may not reflect the ideas of the whole group. Specifically, these findings speak to the importance of the teacher carefully structuring and facilitating students' collaborative concept mapping. For example, our examination of transcripts revealed that the teacher focused more on the procedural aspects of map construction, than on emphasizing deep conceptual connections. This may have affected students' engagement with their map construction. Perhaps map quality might improve if teachers explicitly emphasize the importance of making deep conceptual connections in student discourse related to their maps, help students to establish group collaboration norms, and promote metacognitive reflection. Finally, our study sheds light on forming groups for classroom collaborative activities. Our findings lend support for the effectiveness of heterogeneous group composition to promote negotiation of divergent perspectives towards a shared understanding.

## Directions for Future Research

Future research could systematically investigate collaborative concept mapping with more groups in the classroom. Utilizing both audio and video to match student dialogue to the shared map referent may assist in this kind of analysis. Novak & Cañas, (2008) argue for presenting a *focus question* to facilitate richer map construction. Future research could explore both the effectiveness of focus questions in generating student negotiation towards shared knowledge construction and their impact during collaborative concept mapping. The growing interest in collaborative concept mapping and other collaborative activities in the classroom emphasizes the need for more research to understand the interactions occurring during collaboration and how these impact student learning.

## References

Basque, J., & Lavoie, M.C. (2006). Collaborative concept mapping in education: Major research trends. In A.J. Cañas & J.D. Novak (Eds.), *Concept Maps: Theory, Methodology, Technology. Proceedings of the Second International Conference on Concept Mapping.* San José, Costa Rica.

Daley, B.J., Conceição, S., Mina, L., Altman, B.A., Baldor, M., & Brown, J. (2008). Advancing concept amp research: A review of the 2004 and 2006 CMC research. In A.J. Cañas, P. Reiska, M. Åhlberg, & J.D. Novak (Eds.), *Concept Mapping: Connecting Educators. Proceedings of the Third International Conference on Concept Mapping.* Tallinn, Estonia, & Helsinki, Finland.

Edmondson, K.M. (2000). Assessing science understanding through concept maps. In J.J. Mintzes, J.H. Wandersee, & J.D. Novak (Eds.) *Assessing science understanding: a human constructivist view* (pp. 19-40). San Diego, CA: Academic Press.

Fischer, K.M. (2000). Overview of knowledge mapping. In K.M. Fischer, J.H. Wandersee, & D.E. Moody (Eds.) *Mapping Biology Knowledge* (pp. 5-23). Dorderecht, The Netherlands: Kluwer Academic Publishers.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76*(3), 413-448.

Novak, J. D. & A. J. Cañas, The theory underlying concept maps and how to construct and use them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008, available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf

Puntambekar, S., Stylianou, A., & Goldstein, J. (2007). Comparing classroom enactments of an inquiry curriculum: Lessons learned from two teachers. *The Journal of the Learning Sciences, 16*(1), 81-130.

Roth, W. M., & Roychoudhury, A. (1994). Physics students' epistemologies and views about knowing and learning. *Journal of Research in Science Teaching, 31*(1), 5-30.

Ruiz Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33,* 569–600.

Rummel, N., & Hmelo-Silver, C. (2008). Using contrasting cases to relate collaborative processes and outcomes in CSCL. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F. Prins, (Eds.), International perspectives of the learning sciences: Cre8ing a learning world. Proceedings of the Eighth International Conference of the Learning Sciences (ICLS 2008), Vol 3 (pp. 346-353). International Society of the Learning Sciences, Inc.

Schwartz, D. L. (1999). The productive agency that drives collaborative learning. In: P. Dillenbourg (Ed.), Collaborative learning: Cognitive and computational approaches. (pp. 197–219). Amsterdam.

Stahl, G. (2004) Building Collaborative Knowing- Elements of a Social Theory of Learning, in Strijbos, Kirschner, and Martens (eds.) *What We Know About CSCL in Higher Education*, Dordrecht: Kluwer, 53-86.

Sharples, M. (1994). Computer support for the rhythms of writing. *Computers and Composition, 11,* 217–226.

Teasley, S.D., & Fischer, F. (2008) Cognitive Convergence in Collaborative Learning. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F. Prins, (Eds.), International perspectives of the learning sciences: Cre8ing a learning world. Proceedings of the Eighth International Conference of the Learning Sciences (ICLS 2008), Vol 3 (pp. 354-359). International Society of the Learning Sciences, Inc.

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning & Instruction*, 17(4), 416-426.

van Boxtel, C., van der Linden, J., Roelofs, E., & Erkens, G. (2002). Collaborative concept mapping: Provoking and supporting meaningful discourse. *Theory into Practice, 41*(1), 40-46.

## Acknowledgements

The research described in this paper is supported by an IERI grant (#0437660) from the National Science Foundation to the fourth author.

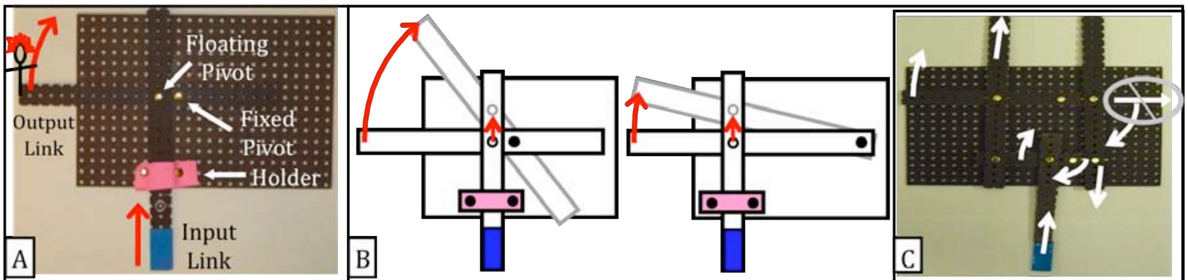
# Embodied Experiences within an Engineering Curriculum

Molly Bolger, Marta Kobiela, Paul Weinberg, Rich Lehrer, Vanderbilt University,
Department of Teaching and Learning
0230 GPC, 230 Appleton Place, Nashville, TN 37203-5721
Email: molly.bolger@vanderbilt.edu, rich.lehrer@vanderbilt.edu

**Abstract:** Although simple mechanisms are commonplace, reasoning about how they work—mechanistic reasoning—is often challenging. To foster mechanistic reasoning, we engaged students in the third- and sixth-grades in the design of kinetic toys that consisted of systems of linked levers. To make the workings of these systems more visible, students participated in forms of activity that we conjectured would afford bodily experience of some of the properties of these mechanisms: constraint and rotary motion. Students progressively re-described and inscribed these embodied experiences as mathematical systems. We report a microgenetic study of one case study student, tracing how embodying and mathematizing motion supported the development of reasoning about how levered systems work.

## Introduction

Reasoning about mechanisms (i.e., mechanistic reasoning) is a key to understanding the designed world. Yet, despite the ubiquity of machines in our culture, and concerted efforts to support mechanistic reasoning in schooling (e.g., simple machine curricula in elementary grades), most students continue to find this form of explanation challenging (Lehrer & Schauble, 1998; Metz, 1991). To support the growth of mechanistic reasoning, we engaged elementary students in the design and construction of kinetic toys. However, our approach to learning by design included efforts to embody and mathematize (Kline, 1980) mechanical systems, such as the simple levered machines depicted in Figure 1. Though mathematical ideas are an intricate part of explanations for mechanical systems, engineering curricula for children often neglect mathematics as a resource (Prevost & Nathan, 2009). Our guiding conjecture was that mathematical description of levered systems would help children reason causally about objects and relations within the system. Our approach was anchored in a body-syntonic (Papert, 1980) wherein initial mathematization emerged from bodily activity, which was then re-expressed with the operation of the simple machine. The instructional design was informed by a prior study of children's naïve reasoning about these simple machines, in which children predicted and explained the direction and amount of output motion.




A: Levered machines are constructed from pegboard and links. Floating pivots are formed by connecting two links and fixed pivots by connecting a link to the pegboard. Holders restricted a link's sideways motion. Outputs were marked with stick figurines.

B: Key: Floating Pivot ○ Fixed Pivot ● Amount of Input Push ↑

Contrasting machines A1 and A3 were used to elicit children's reasoning about relative amounts of motion for input and output. Given the same input push, the output for Machine A1 will move much further than that for Machine A3.

C: Arrows show motions that could be visualized when reasoning about motion of this machine. The circled and crossed out arrow shows impossible motion that preserves the static shape of the machine.

Figure 1. Pegboard machines explained by children.

## Children's naïve mechanistic reasoning

Children in our previous study (Bolger, Kobiela, Weinberg, & Lehrer, 2009) exhibited a wide range of ability in predicting and explaining machine motion. Mechanistic explanations existed, but more commonly children explained machine motion by describing pattern regularities that they noticed (ie they related a cause and effect without explaining the mechanism for the relationship) or they provided fragmented explanations with some of the elements necessary for a mechanistic explanation. Some aspects of levered machine's mechanisms seemed particularly problematic for most children. First, most did not seem to "see" rotation in ways consistent with the operation of the machines. For example, children's gestures often indicated impossible motions that preserved the shape of the machines (Figure 1C). Talk or gesture to indicate rotation of links was not common in most children's explanations. Among the 9 children interviewed, 3 never indicated rotation and 4 did so only in one or two instances. In 22 instances where children drew paths with their fingers to predict output motion that should have been arced, 17 drew straight lines. Even after children moved a machine, they infrequently described rotary paths, seemingly focusing on starting and ending points rather than intermediate motion, as described by Piaget, Inhelder & Szeminska (1960). Second, most children rarely if ever suggested that fixing a link to the board would constrain its motion (4 children never did so and 3 did so on one occasion). Out of 72 total machines explained, there were only 6 instances in which a child recognized *both* constraint of the fixed pivot and rotation, emphasizing the difficulty of *coordinating* these two ideas. Third, children also had difficulty reasoning about relative amounts of input and output motion (which is important to understanding the concept of leverage). Children rarely paid attention to how far links moved, even when a paired contrast was used to draw their attention to this feature.

## Theoretical framework for designing instructional supports

In summary, students had difficulties with: 1) Recognizing an output link's rotation and causally ascribing that rotation to the constraint of the fixed pivot; and 2) Noticing and reasoning about the amount of motion traveled by an output link. In an effort to support these foundations of mechanistic reasoning we designed instruction to support development of reasoning about constraint, rotation and amount of relative motion of inputs and outputs.

Our instructional design is centered on embodied experiences that support student understandings of the mechanisms of the physical system and that provide a prospective pathway for mathematization of the system. Figure 2 suggests that mechanistic reasoning emerges from coordination of embodied experience and mathematization, a perspective consistent with Papert's (1980) conjecture about the role of embodied experience in mathematical development and the more recent work of Lakoff and Nunez (2001) and Barsalou (2008).

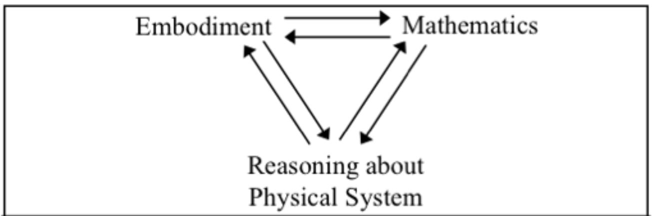



Figure 2. Supporting Learning.

Coordination between embodied experiences and mathematics is consistent with Freudenthal's (1973) notion of "progressive mathematization." Freudenthal states that to implement a program of progressive mathematization successfully, students have to use their experience (e.g., embodied experiences) and invent mathematics through well-developed instructional activities. We drew upon several decades of work leveraging experiences of walking (e.g., a path perspective) as grounds for developing a mathematics of space (e.g., Jordan & Henderson, 1995; Lehrer, Randle, & Sancilio, 1989). In this study, we capitalized on the students' embodied experiences through instruction that supported incorporation of experiential relations into more formal mathematics. These experiences allowed for students to take a path perspective of a moving part within the levered system. Further, we sought to help students integrate these mathematical ideas into their developing explanations for the function of the levered machines.

## Disruption of "Straight": The Embodiment of Rotation

This section describes one embodied task designed to take advantage of the relations in our theoretical model (another task was also designed, but we do not describe it here). We unpack task features that could directly contribute to multi-modal conceptual development (Barsalou, 2008) and highlight the relevant mathematical features of the system as potentially experienced by a student. This task was framed around an apparent tension or disruption, thereby problematizing the construction of an explanation for the experienced phenomenon.

The embodied task (illustrated in Figure 3A) was designed to address two qualities of the mechanisms: 1) fixing a link in one place generates a circular path and 2) the length of this circular path is related to the distance from the pivot to the output. We first engaged students in a conception of "straight" from a path perspective as no turns—a constant heading while walking. We then paired students - one held one end of the rope and remained in place (acting as the fixed pivot) while the other held the other end of the rope and attempted to walk in a straight path perpendicular to the rope's orientation (See Figure 3A). The student walking literally experienced the effect of constraint on path, while the student in the "center" experienced the force required to accelerate (constantly changing direction) the walker. The task was experienced twice with different lengths of rope. The constraint disrupted straightness and produced a circular (i.e., rotary) path, turning a constant amount as the student continued to walk at about the same rate. Ideally, this disrupted straight path, and its description as a mathematical object, could serve as a resource for explaining the operation of a system of levers.

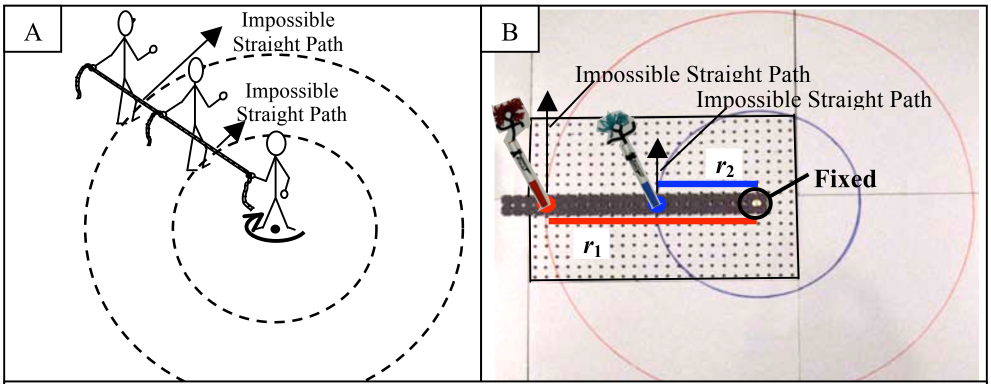



Figure 3. Disrupting "Straight" through Embodied Circular Walks with Rope.

Table 1 elaborates the task's embodied experiences, the mathematics involved in those experiences, and their relations to the physical levered system. During the embodied task, the student walking experienced the disrupted straight path: as he tried to walk straight, the constraint of the rope forced him to turn. Later when the student would be asked to reason about actual levered machines, the system's constraints would be made real by the experience of feeling the rope's pull. To help bridge between the embodied experiences and the physical system "little men" figurines were placed on the links to resemble walkers within the system (Figure 3B). The student who stood in the center, holding the rope, experienced standing in place, feeling "stuck," while pulling the other student toward a circular path. For her, the role of the fixed pivot was magnified as important in resisting the other student's pull to move straight, a resource as she later reasoned about the role of the fixed pivot in the physical system. The different perspectives taken by the two students could influence how the experience served as a resource for reasoning about the physical system. The perspective of the first student was that of a circular path, seeing himself move *around* the student holding the other end of the rope. As he walked, he may have also felt his body constantly turning. When the rope was shortened, this turning accelerated and he may have felt himself going around faster (that is greater angular rotation in the same amount of time) because of the decreased path length. The perspective of the second student did not include a circular walk, but seeing the walker move around her, furthered her experience as "center." When the rope was shortened, the she may have seen the other student moving around her at a faster rate. We conjecture these perspectives are significant in helping students mentally animate the operation of the pegboard machine.

The mathematization of the embodied experience is intended to help highlight relevant features of the experience and the relations between those features that can then be used to reason about the physical system. The stationary student may be thought of as a center of a circle, where the circle is created by the person walking along the constrained path caused by the fixed length of the rope. The rope gives significance to the (otherwise invisible) constant radius of the circle. In this sense, the mathematical features of the experience map directly onto the physical system: the person in center of the circle becomes the fixed pivot, the walking student becomes an output on a link traveling in a circular path. Additionally, the mathematical features can be used to help explain the "unable to walk straight" phenomenon (the path *must* be circular *because* there is a constant radius, - $r_1$ and $r_2$ in Figure 3B). Because a radius exists between the fixed pivot and the output, this mathematical relationship causally connects the constraint of the fixed pivot with the rotary motion of the link. The shorter rope may be then thought of as a shorter radius ($r_2$). In this sense, the path walked is shorter because the circle created is smaller. Mathematically, this occurs because of the proportional relation between the radius and the circumference. Later, when students are asked to reason about the *amount* of motion of an output as it relates to the distance between output and fixed pivot (i.e., radius), they can take the perspective of walking a smaller or larger circle using the longer or shorter rope.

Table 1: Embodied Experiences, Mathematics and the Physical System Related to the Rope Walk.

| Embodied Experience | Mathematics | Physical System |
|---|---|---|
| A ***person*** holds one end of a ***rope***. Another ***person*** holds the other end. | A point (***center***) is drawn with a line (***radius***) ($r_1$) extending out from it. Another ***point*** is drawn at the other end. | A ***link*** is attached to a pegboard at the end with a ***fixed pivot***. The other end is marked as the ***output*** (with a figurine). |
| One person tries to walk straight while the other rotates in place. His straight path is quickly disrupted by the constraining force of the rope held by the other person. | One end of the radius cannot be swept in a straight path without moving *both* ends of the line. | The output cannot be moved in a straight line without removing the fixed pivot. |
| The person walking is forced by the rope onto a ***path going around the person*** rotating in place. The walker is always the ***same distance*** from the center person. | A ***circle*** is created by sweeping the radius (***constant length***) ***around the center point***. | The output ***rotates around the fixed pivot***. At any point in the rotation, the output always remains the ***same distance*** from the fixed pivot. |
| The rope is now exchanged for a ***shorter rope*** and the person walking again attempts to walk straight. | The new, ***shorter, radius*** ($r_2$) is created. Again, the radius cannot be swept in a straight path without moving both ends of the line. | The output (figurine) is now moved in ***closer to the fixed pivot***. Again, the output cannot be moved in a straight line without removing the fixed pivot. |
| The walker is again forced onto a path going around the person rotating in place. This time, the path is ***shorter*** and the walker goes around ***faster***. | A new ***smaller circle*** (with smaller circumference) is created when the shorter radius is swept around the center point. | The output again rotates around the fixed pivot. This time, the path looks ***smaller***. |

## Methods

A microgenetic method was employed, featuring one child and one teacher, lasting an average of approximately 7.2 hours over 8 days. The rope embodiment activity, described above, took place on the first day of instruction. Around the fourth day of instruction, most students participated in a second embodiment activity designed to highlight relative input and output motions. Most instruction was centered around simple design challenges (for example, 'Can you make something that moves using these 3 links?' or 'Can you change this machine so that the output moves in the opposite direction?'). Students also designed and constructed a MechAnimation (a decorated toy driven by levers). Participants (n=11, 5 male) (5 third graders, 6 sixth graders) attended two urban schools in the southeastern United States. Sixty-percent of students at the middle school and ninety-percent at the elementary qualified for free or reduced lunch. The children were ethnically diverse and represented a wide spectrum of achievement in school. To assess gains in predicting and explaining machine motion, children were interviewed before and after instruction.

We chose to carefully study the learning of one child, Sarah. We chose Sarah primarily because of her willingness to express her thinking to us. Though she was a fairly successful student in school, she did not demonstrate a strong naïve ability to predict or explain levered machines in her pre-interview. Analysis of Sarah's learning included coding of her interviews (using the framework from our first study), group video noticings (Jordan & Henderson, 1995) of each day of instruction, and microgenetic analysis of the transcript and video to trace prospective relations between embodiment experiences and other instructional activities.

## One Student's Use of Embodied Activity in Reasoning about Pegboard Machines

We follow Sarah as she engaged in the embodiment task, represented her experience through drawing, and considered the operation of a single link connected to a pegboard. We highlight ways that the experience seemed to be a resource for reasoning about difficult concepts—that is, how constraint of the fixed pivot necessitates rotary motion, mental animation of the system and relations influencing amount of lever motion.

### Sarah's Embodied Experience

As Sarah performed the embodiment task, described earlier, the teacher took hold of one end of the rope and asked Sarah to hold another part of the rope several feet away. The teacher then asked Sarah to "try walking in a straight line that way" (pointing to indicate direction) while keeping the rope taut. Sarah immediately sensed difficulty, at first hesitating ("but uh") and then asked whether she should "turn." The teacher repeated the request to "try to walk in a straight line." As Sarah walked, she described feeling constrained and then *having* to turn. Note that italics in transcript represent ideas that we wish to emphasize; student's or teacher's emphasis is denoted by all caps.

> S: The um-*I can't move any more*…Like the rope can't (?) you *have to turn*.
> T: And then if you kept on just moving (Sarah continues the rotary walk)…what would start to happen?
> S: I would go in like (gestures in a circle with her arm) a *circle*.

Feeling the force of the rope constraining her, Sarah acknowledged that her path was circular. By taking this perspective, she seemed to experience a causal relation between the constraint of the rope and the rotary path. The teacher then highlighted the constant rope length of the walk. She asked Sarah to approximate the amount of rope between them, first standing at one position ("a fourth of the rope") and then standing at another position ("*still* a fourth"). The teacher asked about *all* possible positions: "And then so *anywhere* we turn?" Sarah completed the generalization by concluding that "it's *still* a fourth," By highlighting this feature of their experience, the teacher attempted to help Sarah relate the circular path of her walk to the unchanging distance between her and the teacher.

Finally, Sarah was asked to walk the circular path again, but with a shorter rope length. This time, Sarah was not directed to walk straight but instead immediately asked to "turn." While walking, Sarah initially observed that the path was again a circle and when the teacher asked her for more details, she noticed that "it'd be smaller." The teacher again scaffolded a generalization of the experience, this time emphasizing the relation between rope length and circle size. Moving her hand along the rope, she asked, "so when this becomes smaller?" Sarah completed the generalization, "the circle gets smaller." Thus, through the contrasting walks, a sense of magnitude as "smaller circle" emerged, as well as a relation between the rope length and the distance of the circular path.

## Inscribing the Embodied Experience: Moving towards Mathematization

Immediately following the embodied activity, the teacher requested that the student draw two pictures to show the two different rotations: with the long and the short rope. Her drawing (see top of Figure 4A) showed two circles, one smaller than the other, each with a line extending from it with a point at each end. As Sarah began to explain the drawing, it was evident that she had mathematized all of the relevant *components* of the experience (that is, the people as points, the rope as a line and the walk as a circle), but not all of the *relations* between them. Although her drawing accurately depicted relative differences in circle size and radius (smaller rope, smaller circle), it did not represent radius as moving *around* a center and the circular path as emerging from that interaction. However, Sarah realized this problem when attempting to reenact the experience within the drawing.

> S: I drew me (points to top dot in her drawing) and then you (points to bottom dot in her drawing) and then (moves finger back to top dot) I was holding the rope and (moves finger back to the bottom dot) then *when*, no but that's wrong.

Sarah's inability to animate her drawing motivated a new one (see bottom of Figure 4A). This time, she animated the static representation as she drew and explained. As the static components of the drawing were made dynamic, relations among them emerged. While explaining her drawing, she not only imparted agency within it, but also re-enacted the experience (via gesture) through it: "This one (points to the dot inside the rightmost circle) would be you. And me (points to the dot on the circle) holding the rope *going around* (moving her finger around the circle)."

Sarah's first drawing suggests that while she related "rope length" to circle circumference, this length had not yet been fully mapped to "radius". As she reasoned about the embodied experience and the representation together, her mathematization of the system showed structural relations of the circle—the rope became radius and the teacher (precursor for the fixed pivot) moved *inside* the circle. Her final drawing also symbolized three ideas that became important later in the session – fixed pivot as the center of the circle, the constant radius of the circle, and the relationship between radius and circumference.

## Reasoning about Rotation around a Pivot: Mapping to the Physical System

Within the pegboard system, Sarah's notion of "center" began to take on functional significance. After inscribing the embodied experience, the teacher gave Sarah a pegboard, a link, and a brad fastener and asked her to attach the link to the pegboard any way she wanted. Sarah placed the pivot at the very end of the link, a structure resembling the teacher standing and holding the end of the rope. Interestingly, when asked to describe her machine, Sarah spontaneously mapped the link's movement to this experience: "it [the link] can move around but, yeah like it can move around like rotating *how we did*." Moreover, as she continued describing the machine, she noted new relations among the parts of the system. We suspect that the literal resemblance of the machine's structure to the embodied experience perhaps afforded a "bird's eye" perspective of her previous experience that made these relations salient.

> S: …it *can only rotate* in like this spot (points to the fixed pivot) and no other (waves hand over the mechanism) spot…Like the starting point (touches fixed pivot) you can only start here (touches pivot again) and not start (points to the left of the fixed pivot) somewhere else because this (points to the fixed pivot multiple times) is *stuck* to THAT (points to pivot again) place.

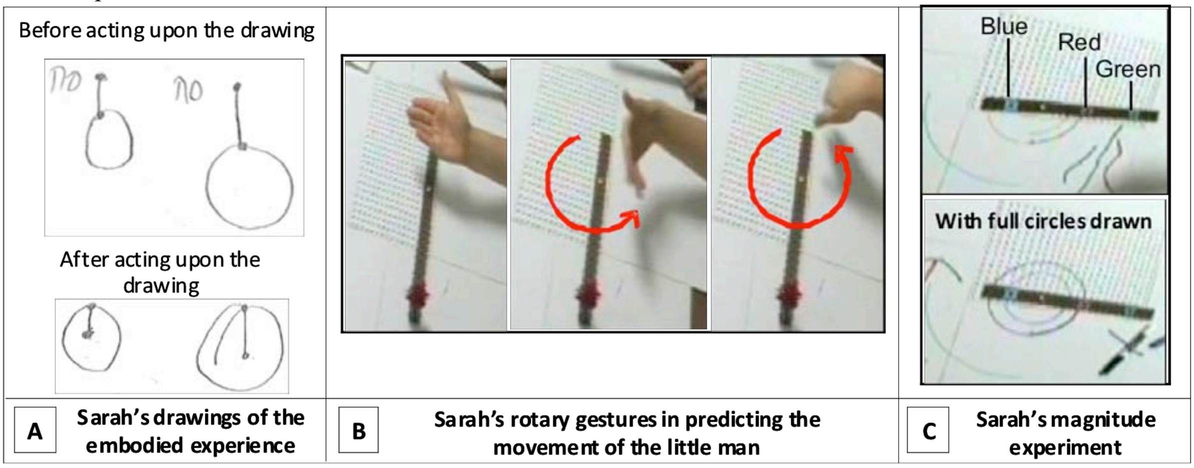



Figure 4. Sarah's use of her embodied experience.

Sarah's notion of "stuckness" attributed to the fixed pivot illustrates an important notion of constraint: that it is unable to move. Furthermore, by describing the rotation as only occurring "in like this spot," she indicated some sense of the source of rotation originating from the fixed pivot.

These ideas of coordinating the rotation of the link to the constraint of the fixed pivot resurfaced again a little later in the session. The teacher had moved the fixed pivot in towards the center of the link and had placed a little man figurine on the left end. She asked Sarah to predict the motion of the little man supposing they were to push the link, pointing to a particular place to the right of the fixed pivot. Sarah correctly predicted the direction of the man, invoking her knowledge of the link's rotary motion. She explained that, "this [the link] will *spin*," gesturing with her hand in a circular motion (Figure 4B). During her pre-interview, Sarah didn't talk about turning or spinning and didn't make rotary hand gestures (with the exception of her noticing the rotary motion of a round-shaped piece of pegboard), suggesting a different way of thinking that may have been supported by the tasks that proceeded this episode.

After testing Sarah's prediction, the teacher asked her how she had known that the path would be curved. In her explanation, Sarah again invoked the importance of the fixed pivot's immobility, but this time indirectly referencing the experience of trying to walk straight: "…it's not going up, it's just … on one brad and when you hit it (gestures pushing link), it will go, that way (gestures rotation) cause it (points near the fixed pivot) *can't always go straight*. Yeah cause it can't move up. It's *in its place*." Here, unlike in her earlier talk, the rotation of the link and the "stuckness" of the fixed pivot are more than simply related. Rather, Sarah now attributes the rotation to the constraint of the fixed pivot, giving it functional significance.

## Using Mathematics of the Circle to Reason about Magnitude of Link Motion

The final task undertaken on the first day of instruction was to focus on the relationship between two distances - the distance from a point on the link to the fixed pivot (radius) and the distance traveled by that point (arc length or circumference). This relationship forms the precursor for understanding the relationship between the magnitude of input and output motions. For this task, the teacher stipulated that the link would be rotated ¼ turn, marking the starting and stopping points. She then asked Sarah to choose 3 places on the link to put colored stickers to correspond with where the "little men" would be placed (Figure 4C, top image).When asked to predict how each sticker would move, Sarah understood that all stickers would rotate, but their paths would have different lengths: "I think that. They will move different-No they won't move differently but they're um (points to the board and then moves finger in a circle) lengths I think will be different." When the teacher talked further to Sarah about her prediction, Sarah spontaneously referred to her embodied experience "So they'll like (gestures finger in circles again) turn the same way but like our *rope* was different lengths and their shape is going to be smaller or bigger." Hence, Sarah seemed to use her experiences to reason that the rope length (radius) would affect the size of path walked (circumference) in this simple physical system. However, her specific mapping of the "rope length" onto the link was less clear as she went on to incorrectly predict the relative distances that each sticker would travel.

Sarah was guided by her teacher to find a generalization that would allow her to accurately predict the relative amounts of motion for points along the link, but she eventually came to an impasse. She explained her first prediction (blue>green>red) this way: "So yeah like the outer it is the bigger it's gonna be so this (points to blue sticker) is outer". Here, Sarah would be correct, if the link rotated around its center rather than the fixed pivot. After measuring the relative amounts traveled by the stickers (green>red>blue), Sarah created a new incorrect explanation to fit the observations at hand, "I think that the closer it is to (points to green sticker) this end, it will go the most and the farthest it is to that end (points to the right end of the link), it will go the least." When the teacher offered a counter example to disprove this new theory (the distance traveled by a point at the far left end is greater than the distance traveled by the blue sticker), Sarah was unable to pose a new explanation. At this point, it seemed that the fixed pivot was invisible when Sarah reasoned about these relative distances.

Finally, the teacher helped Sarah draw the full circles traveled by the stickers and by the point at the far left end of the link (Figure 4C, bottom image). This far left point, which began as an unexplainable counter example, was repurposed to provoke an explanation that involved the radius and the fixed pivot as center of the circle:

> T: So why do you think the [point at the left end] and the red went the same? Are on the same *circle*?...
> S: (gasps) Oh:: Maybe because like uh this one is (points to the far left end) one two three four (counting the holes) five six and that one's (points to the red sticker) one two three four five six. It's the same distance close to the brad.

The relevant mathematical relationship seemed to appear when the circles were drawn. This teacher move served a dual purpose – drawing the visual focus to the fixed pivot as center and bringing the task closer to the embodied experience. In the preceding section, we saw that Sarah seemed to draw from her embodied experience as she reasoned about the direction of turn for this same machine. However, when she moved to a slightly more complex task, seemingly small details (i.e. locating the pivot away from the end of the link and drawing arcs instead of complete circles) may have been enough to disrupt a literal mapping to the embodied experience. This speaks to the importance of teacher sensitivity about appropriate scaffolds that will allow for best use of embodied experiences. Interestingly, on later days of instruction, Sarah spontaneously invoked the embodied experience when asked to reason again about the radius-circumference relationship, suggesting that the experience was a useful support beyond its immediate context. For example, in the third day of instruction, when Sarah was asked again to explain the same phenomenon (also with a single link mechanism), she provided the following explanation.

> T: Why is it when it's really far it's making a big circle, like what MAKES the circle?
> S: Cause us like when we were doing the *rope*. There is only like, pretend this is me (points to green sticker) and that was you (points to pivot). There was only like this much *distance* between us.
> T: You can only like (turns her body in constrained manner)
> S: Yeah, like move that much.

Sarah's difficulty with reasoning about magnitude of link motion prompted the development of another embodied experience used later in instruction.

## Conclusion

We designed embodied experiences (Papert, 1980; Abelson & diSessa, 1980) to support "mathematization" (Kline, 1980) of mechanical systems, namely those containing simple levered machines. Drawing from our previous studies, we targeted particular naïve student conceptual difficulties that were seen in most students, even when working with the machines. These were: viewing rotary motion in the machines, attending to the amount of output link motion, and reasoning about the constraint resultant from fixing a link to the pegboard. Embodied experiences served as a resource for progressive mathematization of the system (Kline, 1980; Freudenthal, 1973), aided by strategic teaching supports, such as: asking the student to represent the literal embodied experience on paper, asking the student to reason about a physical system that closely resembled the embodied experience and frequently revisiting how relevant features of the embodied experience mapped to the physical system.

With our case study student, Sarah, the embodied rope experience and subsequent teaching supports seemed to help her reason about each of our target concepts. Though Sarah did not address the rotation of links in her pre-interview, this idea flowed readily from the embodiment experience. Within the experience, she noticed herself going "around," an idea later resurfacing in her drawings, talk and gesture. Though this experience helped Sarah readily see circular paths, she required additional supports to see other mathematical features of circles. Reasoning around her representation seemed to help Sarah map the "rope" to radius. As Sarah revised her drawing, she appeared to reenact the experience, imparting agency and replaying motion through gesture. Further, the relationship between radius and circumference, a necessary precursor to understanding the *amount* of output motion, was not obvious to Sarah. By helping Sarah work with the physical system (and adjusting it to visually cue the embodied experience), the teacher was able to guide her to discover this relationship and map it in a lasting way to the embodied experience.

Sarah's process of reasoning about how the fixed pivot served to constrain motion was refined throughout the first day of instruction. During the embodied experience, she talked about how her walk was constrained to a circular path, but the source of this constraint was not evident. As she refined her representation, the fixed pivot (teacher) was moved *inside* the circle. When she built the simple physical system and mapped it back to the embodied rope experience, she noted for the first time that the fixed pivot was "stuck" and assigned it significance as the source of rotation. When directed to reason about the path of a "little man" on the link she causally connected his circular motion with the fixed pivot as constraint. However, as instruction continued Sarah had difficulty locating the fixed pivot as the circles' center as well as maintaining a notion of fixed pivot as the source of the disruption of straight. We conjecture that Sarah's lack of experience acting as the fixed pivot in the embodied experience may have made her less aware of its role as a constraint.

Our results suggest that integrating mathematics into an engineering curriculum can support students as they begin to construct explanations for simple machines. However, many questions remain as to how best integrate mathematics in ways that will be accessible and meaningful to students and that will best support mechanistic reasoning. Current research is focused on answering these questions through further microgenetic analysis as well as analysis of data from intact classrooms. Classroom instruction used similar embodied activities, but was able to capitalize on the variety of student ideas. Comparison of ideas was used to motivate student reasoning.

## References


Abelson, H. & diSessa, A.A. (1980). *Turtle Geometry: The Computer as a Medium for Exploring Mathematics.* Cambridge, MA: The M.I.T. Press.

Barsalou, L.W. (2008). Grounded cognition. *The Annual Review of Psychology, 59,* 617-645.

Bolger, M. Kobiela, M., Weinberg, P., Lehrer, R. (2009). Analysis of Children's Mechanistic Reasoning about Linkages and Levers in the Context of Engineering Design. *2009 Conference Proceedings, American Society for Engineering Education*, AC 2009-1279.

Freudenthal, H. (1973). *Mathematics as an Educational Task.* Dordrecht, Holland: D. Reidel Publishing Company.

Jordan, B. and Henderson, A. (1995) Interaction analysis: Foundations and practice. *Journal of the Learning Sciences 4*(1), 39–103.

Kline, M. (1980). *Mathematics. The loss of certainty*. Oxford, UK: Oxford University Press.

Lakoff, G., & Nunez, R. (2001). *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics into being.* New York, NY: Basic Books.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41(3),* 635-679.

Lehrer, R., Randle, L., & Sancilio, L. (1989). Learning preproof geometry with LOGO, *Cognition and Instruction, 6(2),* 159-184.

Lehrer, R. & Schauble, L. (1998). Reasoning about structure and function: Children's conceptions of gears. *Journal of Research in Science Teaching, 35(1),* 3-25.

Lehrer, R., Schauble, L., Strom, D., & Pigge, M. (2001). Similarity of form and substance: Modeling material kind. In Sharon M. Carver and David Klahr (Eds.), *Cognition and Instruction: Twenty-five years of progress.* (pp. 39-74). Mahwah, NJ: Lawrence Erlbaum Associates.

Metz, K.E. (1991). Development of explanation: Incremental and fundamental change in children's physics knowledge, *Journal of Research in Science Teaching, 28(9),* 785-797.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas.* New York, NY: Basic Books.

Piaget, J. Inhelder, B., & Szeminska, S. (1960). *The Child's Conception of Geometry*. New York, NY: Basic Books.

Prevost, A., Nathan, M. (2009). The Integration of Mathematics into Precollege Engineering: The Search for Explicit Connections. *2009 Conference Proceedings, American Society for Engineering Education*, AC 2009-1790.

# "Romantic" beats "classic": New insights on the effects of self-regulation on learning by writing

Isabel Braun, Susanne Philippi, Matthias Nückles
University of Freiburg, Rempartstr. 11, 79098 Freiburg, Germany
E-Mail: isabel.braun@ezw.uni-freiburg.de, matthias.nueckles@ezw.uni-freiburg.de

**Abstract:** For several decades writing has been advocated as an academic activity that can facilitate knowledge acquisition. Theories of writing and models of writing development that researchers draw on to explain the positive effects of writing to learn emphasize the importance of self-regulation, particularly, planning, monitoring and revision. Yet, there is some empirical evidence indicating that the self-regulation of writing may be detrimental to learning. To analyze systematically the effects of self-regulation on learning by writing we conducted an empirical study. Participants (*n*=87) read an instruction that induced (1) self-regulated writing, (2) spontaneous writing or (3) self-regulated and spontaneous writing. Under the control condition, participants did not receive instruction in writing to learn. All participants then wrote an entry to a learning journal. Results showed that self-regulated writing exerted a negative influence on knowledge construction during writing and on post-test performance. In light of these findings, the emphasis on self-regulation strategies in writing-to-learn instruction should be reconsidered.

## Introduction

As they move through high school and college, students engage in a variety of writing activities. They write to improve their writing and argumentation skills, to document academic work or to demonstrate what they have learned in a class. Teachers also assign writing-to-learn tasks which aim at supporting students in acquiring or deepening their knowledge of subject matter. One of the tenets of research-based writing instruction (e.g., Graham, 2006; Harris & Graham, 1992) is that learners should be taught to self-regulate their writing, i.e. to plan, monitor and revise the rhetoric, structure and content of their texts. However, the rhetorical and structural quality of the written product is not the main target outcome of writing to learn, in contrast to other academic writing activities such as writing a term paper or completing an essay exam. Hence, learners may not benefit from self-regulating their writing when engaging in a writing-to-learn activity. There is some empirical evidence suggesting that adopting a self-regulated stance toward writing, which Galbraith (1992) referred to as "classic", may indeed be detrimental to learning. Galbraith (1992) found that learners who adopted a "romantic" stance toward writing, treating writing as a spontaneous, creative process not involving the planning and monitoring of text characteristics, showed higher gains on measures of knowledge construction than those who adopted a "classic" stance toward writing. However, empirical studies involving a typical writing-to-learn task (such as journal writing) and comparing the effects of the self-regulated ("classic") stance and the spontaneous ("romantic") stance toward writing are missing. Therefore, we conducted an empirical study in which we instructed students to adopt different stances toward writing and had them write an entry to a learning journal. Our aim was to advance research on self-regulation and learning by writing and to contribute to the improvement of writing-to-learn instruction. Before reporting on the study we sketch the theoretical and empirical background relevant to our research.

## Self-regulation in cognitive theories of writing

Despite empirical evidence of the positive effects of writing on learning there are few models of processes of writing to learn. Therefore, researchers draw on cognitive theories of writing to explain the knowledge and achievement effects of academic writing activities. Klein (1999) subsumed research on learning by writing under four sets of theories: Shaping at the point of utterance, forward search, backward search and genre. The perspective on the self-regulation of writing offered by the "shaping at the point of utterance" and "forward search" models differs widely from that of the "backward search" and genre models. As a consequence, they also differ in their implications for writing-to-learn instruction.

The "shaping at the point of utterance" theory was proposed by Britton (1982). The central assumption of Britton's model is that knowledge emerges as ideas are expressed in writing. Therefore, writers should engage in spontaneous writing and not focus on the linguistic and rhetorical characteristics of the written product. That is, they should not self-regulate their writing. The assumption underlying Britton's model forms the backbone of the "forward search" models of writing-to-learn: Producing text in a spontaneous fashion allows the writer to reread and reflect on his ideas which triggers further idea generation, setting of rhetorical goals and text revisions. What Klein has termed forward search is strongly advocated by Galbraith (e.g., 1992). He argues that writing does not yield learning through a sequence of self-regulatory processes (planning, translating, monitoring and revising). Instead, he proposes that ideas are generated as the writer's implicit knowledge and the emerging text interact. Hence, when the goal of writing is knowledge acquisition writing should take the form of "spontaneous spelling-out of ideas in continuous prose" (Galbraith, 1992, p. 45) with subsequent linguistic and rhetorical revisions to the text. Rhetorical goals are not set prior to writing but emerge during writing. Galbraith refers to this stance toward writing as "romantic" – an allusion to romanticist writers' departure from adherence to prescribed rhetorical formats.

Empirical studies on the effects of writing-to-learn instruction that is based on "shaping at the point of utterance" and "forward search" models are scarce. There is some evidence of the benefits of revising notes taken during lectures (Benton, Kiewra, Whitfill & Dennison, 1993) and of rereading and revising entries to a learning journal (Hübner, Nückles & Renkl, 2006). Moreover, studies from writing research indicate that expert writers revise their texts several times to increase the global coherence of their output whereas novice writers make mostly linguistic and rhetorical revisions at the sentence level (Klein, 1999). Theoretical considerations speak in favor of and against instructing writers to adopt a "romantic" stance when engaging in writing-to-learn activities. On the one hand, externalizing ideas in writing may free working memory capacity for deeper processing and the external representation may facilitate the drawing of inferences. On the other hand, the continuous text that is written spontaneously has to be processed in order to trigger idea generation and revisions which may actually increase working memory load (Torrance & Galbraith, 2006).

The "backward search" and genre models provide a different perspective on the self-regulation of writing than the models just described. Among the theories described as "backward search" models by Klein (1999), the Bereiter and Scardamalia (1987) model has been the most influential in writing-to-learn research. According to Bereiter and Scardamalia's *knowledge transforming* model, expert writers move between the content space, which represents their knowledge and beliefs about the topic they intend to write about, and the rhetorical space, which contains their knowledge and beliefs about the context of the writing activity and the rhetorical goals they intend to direct their writing at. As they move continuously between the two spaces, expert writers solve the "problems" of what to say and how to say it by setting rhetorical goals, breaking them down into subgoals, translating knowledge into text to satisfy the goals, checking whether the goals were met and revising the text accordingly. Moving between the two spaces is supposed to yield learning as writers transform, i.e. restructure, their knowledge as they do so. The Bereiter and Scardamalia model is similar to genre models of writing (Klein, 1999) in that self-regulatory processes directed at the production of high-quality text are considered essential to learning. The core assumption of the genre models (e.g., Langer & Applebee, 1987) is that writing genres impose specific structural, linguistic and rhetorical constraints. In order to meet the requirements of the respective writing genre, learners have to self-regulate their writing which in turn induces deep-level cognitive processes, for example, identification of links between arguments and critical evaluation of counterarguments in the case of a compare-and-contrast essay.

Most of the evidence presented in support of the Bereiter and Scardamalia (1987) model stems from studies comparing expert writers' strategies and novice writers' strategies. The studies converge on the finding that expert writers plan and revise their texts more often and at a deeper level than novice writers. And the extent of their planning and revision activities is related to the quality of the texts they produce (for a review, see Klein, 1999). Kirkpatrick and Klein (2009) recently demonstrated that training secondary students to use planning strategies when writing analytic essays increases the quality of the texts written by the students. As with most studies on writing, however, Kirkpatrick and Klein did not include outcome measures of knowledge in their study. There are studies which provide support to the assumption of the genre models. Langer and Applebee (1987), for example, reported on the positive effects of analytic writing on comprehension with secondary students. However, the link between the characteristics of writing genres and the cognitive processes supposed to be triggered by them has not been established.

As pointed out at the beginning of this section, the "shaping at the point of utterance" and "forward search" models and the "backward search" and genre models imply different approaches to learning by writing. Galbraith (1992) contrasted the "romantic" stance toward writing with that of Bereiter and Scardamalia (1987) and termed the latter the "classic" stance toward writing, referring to the strict adherence of the old Greek writers to stylistic conventions. The writer adopting a "classic" stance carefully plans what to write in which way and engages in self-monitoring during writing. In his study, Galbraith provided participants either with a "classic" writing instruction (writers were supposed to plan a text) or with a "romantic" writing instruction (writers were asked to write a text without focusing on its structure and rhetoric). Galbraith found support for the "romantic" approach. Writers who adopted this stance toward writing outperformed those who adopted a "classic" stance on measures of idea generation and knowledge acquisition. However, the participants in his study were not instructed to write in order to learn but to produce a piece of text about a topic of their choice. A controlled experiment including a typical writing-to-learn task is still missing.

Before we report on our study we now turn to research on learning by writing and its implications for writing-to-learn instruction.

## Research on learning by writing

In the 1970s, researchers and educators began to advocate the inclusion of writing activities not only in language classrooms but also in mathematics and science classrooms in high school as well as at the post-secondary level (e.g., Fulwiler & Young, 1982). Although the voices of the *writing-across-the-curriculum* advocates have somewhat faded off their arguments stand strong. Writing has been shown to contribute to knowledge acquisition in a range of academic disciplines (e.g., Applebee, 1984; Connor-Greene, 2000; McCrindle & Christensen, 1995; Rivard, 1994) and to produce positive effects on academic achievement (Bangert-Drowns, Hurley & Wilkinson, 2004). However, the effects identified in writing-to-learn studies are typically small and found under certain conditions only. Nückles, Schwonke, Berthold and Renkl (2004), for example, showed that graduate students who engaged in journal writing over the course of a semester but were not instructed in writing to learn did not demonstrate use of deep-level cognitive strategies in their writing.

Studies such as that by Nückles et al. (2004) indicate that learners have to be provided with instructional support either before or while they engage in writing to learn. Writing research has converged on the findings that self-regulation characterizes skilled writing and determines students' written outcomes (for a review, see Graham & Harris, 2000). Therefore, instructional models of writing development, for example, the *self-regulated strategy development model* by Harris and Graham (1992) put strong emphasis on the teaching of self-regulation strategies, particularly planning, self-monitoring and revision strategies. However, self-regulation may not play the same role in learning by writing and may even be detrimental to learning (cf. Galbraith, 1992). With most academic writing activities (e.g., lab report, term paper, thesis) the texts produced by the students are judged against the standard or conventional characteristics of the respective writing genre. For information on the characteristics of well-written texts, students are expected to draw on the guidelines provided along with the writing task and on the knowledge of text genres which they have acquired through formal instruction and literary socialization. Students' success on a writing-to-learn task, however, is not evaluated based on the rhetorical and structural quality of the written product but based on their achievement on measures of knowledge, performance and academic achievement. Therefore, learners who invest limited cognitive resources (cf. Torrance & Galbraith, 2006) in self-regulating their writing so as to produce texts with certain rhetorical and structural characteristics may not succeed on writing-to-learn tasks.

The assumption that adopting a self-regulated stance toward writing can be detrimental to learning is refuted by the results of Bangert-Drowns et al. (2004). Based on their meta-analysis of school-based writing-to-learn interventions, Bangert-Drowns et al. concluded that metacognitive prompts were among the predictors of learning by writing. However, studies conducted in years following the Bangert-Drowns et al. meta-analysis have shown that facilitating metacognitive self-regulation via prompts may not be sufficient to secure the positive effects of writing to learn. In a study by Berthold, Nückles and Renkl (2007), for example, students were prompted to use cognitive strategies, metacognitive strategies, cognitive and metacognitive strategies or no strategies during writing. It turned out that metacognitive prompts per se did not have a positive effect on learning outcomes. The difference between the studies included in the Bangert-Drowns et al. (2004) meta-analysis and the Berthold et al. (2007) study was the nature of the writing-to-learn task. Berthold et al. had their participants write entries to a learning journal, a writing genre employed for the purpose of learning by writing only and characterized by few rhetorical and structural constraints (Nückles et al., 2004). Bangert-Drowns et al. (2004), however, meta-analyzed studies that included a range of writing assignments. With some writing genres (e.g., analytic essays) the aim of writing to learn - as opposed to learning to write or demonstrating knowledge - may not be salient (cf. Klein, 1999). Metacognitive prompts might take effect with such writing assignments because they direct the writer's attention to learning processes and away from the constraints implied by the writing genre.

Altogether, we argue that instructing students to adopt a self-regulated stance toward writing may not be effective with regard to learning by writing.

## Research questions

As shown in the background sections, cognitive theories of writing and research on learning by writing do not allow firm conclusions as to the effects of self-regulation on learning by writing. As a consequence, it is still unclear which model of the writing process writing-to-learn instruction should be based on – should students be instructed to adopt a "romantic" or a "classic" stance toward writing when asked to engage in a writing-to-learn activity? We addressed this gap in research by conducting an empirical study. In particular, we were interested in the following questions: How do writing-to-learn instructions that differ with regard to the stance toward writing they induce impact on (1) the number and adequacy of ideas generated while writing an entry to a learning journal, (2) the acquisition of content knowledge, and (3) the quality of the written product?

## Method

### Participants and design

The study was advertised through flyers and posters on the campus of a medium-sized research university and involved a sample ($n$=87) of college students ($n$=79) and professionals holding a college degree ($n$=8). With the exception of age, the subsample of professionals did not differ significantly from the subsample of college students on any of the control variables we assessed. The mean age of the total sample was 24.16 (*SD*=4.91). The mean age of the professionals was 27.38 (*SD*=2.87). Females (*n*=59) and males were evenly distributed among the conditions.

A 2x2 experimental design was used with participants randomly assigned to one condition: (1) self-regulated writing (*n*=23), (2) spontaneous writing (*n*=22), (3) self-regulated and spontaneous writing (combined instruction condition, *n*=21), (4) writing without instruction in writing to learn (control condition, *n*=21).

### Procedure and writing-to-learn instructions (experimental variation)

All participants attended a 30-minute video-taped presentation on a topic from the field of educational psychology (cognitive architecture, cognitive load theory and worked-examples effect). They were not allowed to take notes while watching the presentation. Then they were informed about the general aim of writing learning journals (attaining deep-level comprehension and improving strategic learning) and received the writing-to-learn instruction of the respective condition. The spontaneous writing instruction contained three metaphors: Participants were instructed to engage in self-talk, to approach the task with a creative mindset and to treat writing as discovery. Thus, the instruction induced a "romantic" stance toward writing (Galbraith, 1992). In the self-regulated writing condition, the instruction centered on three metaphors. The metaphors were supposed to induce a "classic" stance toward writing (Galbraith, 1992): Participants were instructed to think of writing as rhetorical problem solving, to approach the task with an analytic mindset and to treat writing as a process of continuous reflection and revision. For the combined instruction, we integrated the spontaneous and the goal directed writing instruction. Hence, it contained six metaphors that instigated participants to set and monitor rhetorical goals, write down their ideas as they occurred to them and revise the text if necessary, thus treating writing as a self-regulatory activity and a process of self-talk and discovery at the same time. In the control condition, participants did not receive any instructions beyond the information about the general aim of writing learning journals.

We decided to use metaphors to induce specific stances toward writing for the two reasons. First, persons bound for or holding a college degree can be expected to have at least some experience in writing at an academic level and to have acquired strategies for writing over the course of their academic and professional careers. Therefore, a flow chart or other representation of a writing scheme (as often used in writing instruction) was likely not to be considered by the participants during the actual writing activity because of low perceived usefulness. Second, metaphors are well suited to assist learners in forming a conceptual model of a task as they illustrate complex concepts and trigger associations. Metaphors were successfully used as vehicles for instruction, for example, by Bromme & Stahl (2005) in a study on learning from hypertexts.

### Measures and rating

Participants' knowledge of the presentation topic at post-test (after they had completed the writing-to-learn task) was assessed by means of six free-response items. Three items tested participants' understanding of the concepts directly addressed in the presentation (e.g., "Explain the types of cognitive load and how they interact."). The remaining three items required deep understanding of the presentation topic. Hence, these items addressed transfer of knowledge. One of the items described, for example, a classroom scenario in which students were unable to solve math problems on their own after having been presented one example. Participants were asked to explain the scenario from the perspective of cognitive load theory. We also administered a knowledge pre-test to assess participants' prior knowledge of the presentation topic. The knowledge pre-test consisted of four free-response items: One question on cognitive architecture, two on cognitive load theory and one on worked-out examples. The knowledge tests were graded for the overall quality of the answers on a six-point scale (following Biggs & Collis, 1982) and for the number of correct answers (post-test only). In order to be able to control statistically for other potential confounding variables, participants were asked to provide socio-demographic information and information on their prior experience in writing learning journals. They were also asked to rate the perceived helpfulness of the writing-to-learn instruction.

The texts produced by the participants were analyzed for the number of new ideas, correct statements and incorrect statements they contained. With regard to the quality of the written products, two independent raters assessed the structure and coherence of the texts and the degree to which deep-level cognitive (elaboration) and metacognitive (reflection) learning strategies were evidenced in the texts. Inter-rater reliabilities for these ratings were satisfactory ($ICC_{unjust}$=.71 to $ICC_{unjust}$=.76).

## Results

The participants of the four conditions did not differ significantly with regard to any of the socio-demographic variables, their prior experience in writing learning journals, their performance on the knowledge pre-test and the amount of text they produced. Hence, the subsamples constituting the four conditions were comparable. Across conditions, participants reported low experience in journal writing (number of entries to learning journals written prior to participation in the study: *M*=3.47, *SD*=11.28).

A two-way univariate analysis of variance (ANOVA) with overall quality of the answers on the knowledge posttest as the dependent variable and $\alpha=.05$ revealed a significant main effect for self-regulated writing, $F(1,83)=5.74$, $p=.019$, $\eta^2=.07$ (Figure 1). Further ANOVA analyses also yielded a significant main effect of self-regulated writing on quantitative performance at post-test (number of correct answers on the knowledge post-test), $F(1,83)=7.35$, $p=.008$, $\eta^2=.08$ (Figure 2). Hence, the participants in the self-regulated writing condition and those in the combined instruction condition learned less from writing the entry to the learning journal and acquired a more shallow understanding of the presentation topic than those who received the spontaneous writing instruction or no instruction in writing to learn (control condition).

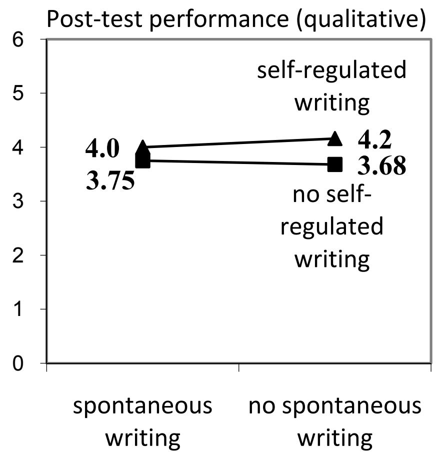



Figure 1. Main effect of self-regulated writing on overall quality of answers on the knowledge post-test (rated on a six-point scale with higher numbers indicating lower quality).

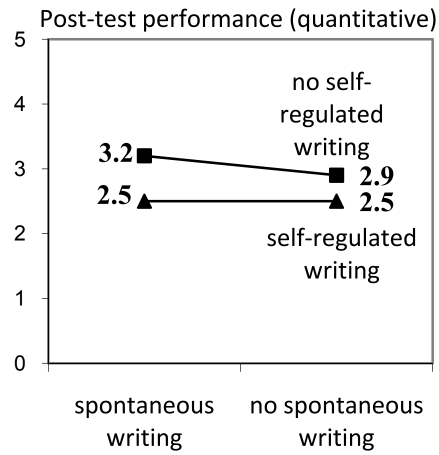



Figure 2. Main effect of self-regulated writing on number of correct answers on the knowledge post-test.

Similar findings emerged from the separate ANOVA analyses we conducted with the number of new ideas and the number of correct statements identified in the learning journals as the dependent variables: The participants who had been instructed to self-regulate their writing produced significantly fewer new ideas ($F[1,83]=4.37$, $p=.040$, $\eta^2=.05$) and fewer correct statements ($F[1,83]=7.41$, $p=.008$, $\eta^2=.08$) than those who had not been instructed to do so (Figures 3 and 4). In all of the four conditions, the average number of incorrect statements made in the learning journals was low. The highest number of incorrect statements was found in the combined instruction condition, followed by the self-regulated writing condition.

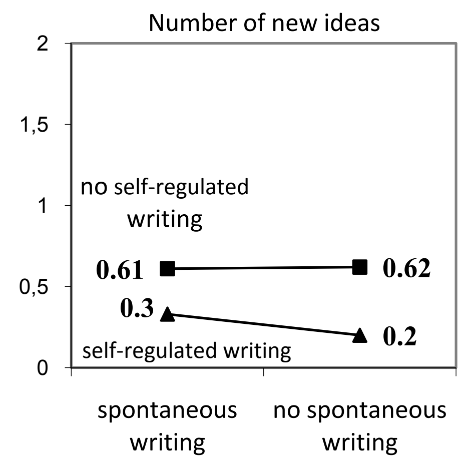



Figure 3. Main effect of self-regulated writing on number of new ideas in the journal entry.

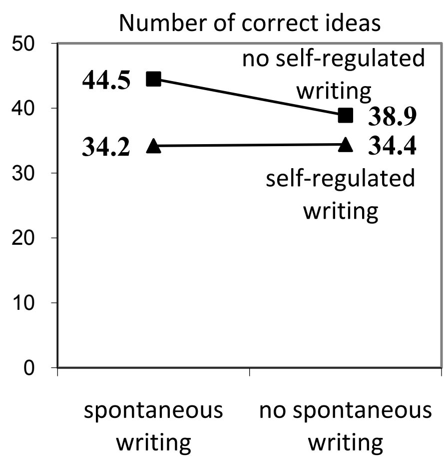



Figure 4. Main effect of self-regulated writing on number of correct ideas documented in the journal entry.

There were no significant differences between the conditions with regard to the quality of the written product (see Table 1 for means and standard deviations). Descriptively, however, participants who had been instructed to adopt a "classic" stance toward writing (self-regulated writing, combined instruction) received lower structure and elaboration ratings than those who wrote spontaneously or without prior instruction.

Table 1: Ratings of the quality of the journal entries written by condition

| | Spontaneous writing condition | Self-regulated writing condition | Combined instruction condition | Control condition |
|---|---|---|---|---|
| Structure *M (SD)* | 3.57 (1.50) | 3.50 (1.53) | 3.33 (1.27) | 3.52 (1.53) |
| Coherence *M (SD)* | 4.04 (1.58) | 4.55 (.96) | 4.00 (1.41) | 4.00 (1.51) |
| Elaboration *M (SD)* | 3.96 (1.49) | 3.82 (1.22) | 3.62 (1.49) | 3.81 (1.25) |
| Metacognition *M (SD)* | 2.04 (1.52) | 2.50 (1.37) | 2.24 (1.64) | 1.48 (1.25) |

*Note.* The journal entries were rated (separately on each dimension) on a six-point scale ranging from 1 (no evidence in the journal entry) to 6 (strong evidence in the journal entry).

Despite the differences in learning processes and outcomes between the four conditions, participants did not perceive the helpfulness of the writing-to-learn instructions to be different. An ANOVA with perceived helpfulness as the dependent variable did not yield a significant result, $F(3,83)=2.37$, $p=.077$.

Altogether, the following picture emerges from the analyses of the knowledge post-test and the journal entries written by the participants: Adopting a self-regulated ("classic") stance toward writing impeded learning (as evidenced by participants' quantitative and qualitative performance on the knowledge post-test). At the process levels of writing and learning, the results are somewhat inconsistent but also point toward negative effects of self-regulation on learning by writing.

## Discussion and future work

The results of our study support the notion that learners should not be instructed to self-regulate their writing when they engage in a writing-to-learn activity. Not only did those writers whom learning by writing had been presented to as a self-regulated activity fare worse with regard to knowledge acquisition. There also was some indication that setting rhetorical goals, followed by planning, monitoring and revision of the emerging text exerted a negative influence on writing and learning processes: Writers who had been instructed to adopt a self-regulated stance toward writing were less successful at demonstrating idea generation, correct knowledge and high-quality writing than those who wrote spontaneously, in a "romantic" fashion.

Our findings corroborate the results of the study by Galbraith (1992) who found that writing spontaneously was superior to careful planning with regard to idea generation and knowledge acquisition. The results of the present study are also in line with recent studies on learning by writing which have called into question the sole effectiveness of metacognitive prompting (e.g., Berthold et al., 2007). Furthermore, the outcomes of the self-regulated writing condition and the combined instruction condition challenge the assumption that findings from research on writing and writing development can be generalized to learning by writing. The fact that the participants in our study were adults whereas those in most studies on writing development were children and adolescents might have played a role. But it is much more reasonable to assume that there are fundamental differences between the processes of writing to learn (and learning how to do so most effectively) and learning to write (cf. Klein, 1999). When designing writing-to-learn instruction this has to be taken into account. An interesting question would be whether changing the focus of the self-regulated writing instruction would have affected the outcomes of our study. In the self-regulated writing condition and in the combined instruction condition, participants were asked to monitor their understanding of the presentation topic but also to set and monitor rhetorical goals. This might have decreased the salience of the actual aim of the writing-to-learn task, i.e. knowledge acquisition. In the future, we plan to conduct a study involving two self-regulated writing conditions: In both conditions writers will be instructed to adopt a self-regulated stance toward writing - but the focus will be rhetoric and structure in one condition and on comprehension in the other. To our knowledge such a design has not been realized with a writing-to-learn task.

At first glance, the outcomes of the control condition do not seem to fit into the general picture that emerged from our study. The participants of the control condition showed strong performance on the knowledge posttest and the quality of their journal entries was almost as high as in the spontaneous writing condition. The apparent contradiction between these findings and studies pointing toward the ineffectiveness of unguided journal writing (e.g., Nückles et al., 2004) may be explained by the very nature of the writing-to-learn instructions. The instruction of the control condition asked participants to write in such a way as to obtain deep understanding but did not refer to any characteristics of the writing genre. This may have lifted the rhetorical and structural constraints off the writing-to-learn task that operated in the self-regulated writing condition and in the combined instruction condition but not in the spontaneous writing condition. Hence, participants in the control condition may have engaged in the same writing behavior and cognitive processes as those in the spontaneous writing condition. This explanation is somewhat speculative though as our design did not include a manipulation check on the experimental variation. Hence, we do not know how the participants processed and implemented the writing-to-learn instructions of the experimental and control conditions. Future studies should consider analyzing the writing process itself (in addition to its product and outcomes) both at the level of writing behavior (e.g., via log-file analysis) and cognitive processes (employing a think-aloud methodology).

Another surprising finding is that we were unable to identify significant differences between conditions regarding the quality of the texts written by the participants. The descriptive differences, however, largely reflect the instructions the participants received in the respective condition. Participants who were instructed to adopt a "classic" stance toward writing demonstrated low cognitive-elaborative strategy use but relatively high metacognitive strategy use. Despite being instructed to self-regulate their writing they received lower ratings for the structure of their texts than the participants who had been instructed to adopt a "romantic" stance toward writing. In the present study, the lack of significant differences and the partially inconsistent results may be explained by the fact that the participants were inexperienced in writing learning journals and wrote a single journal entry only. Studies are needed in which participants write several entries to a learning journal over an extended period of time.

In conclusion, additional studies are required to further clarify the effects of self-regulation on the outcomes (and processes) of writing to learn. In addition to comparing writing-to-learn instructions which direct self-regulation either at text characteristics or at comprehension we suggest that process analyses and long-term interventions be carried out. Along the lines of Klein (1999), we hope to have reopened inquiry into metacognitive processes in writing to learn and sparked further discussion on effective writing-to-learn instruction.

## References

Ackerman, J. M. (1993). The promise of writing to learn. *Written Communication, 10*(3), 334.

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research, 54*, 577-596.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29-58.

Benton, S. L., Kiewra, K. A., Whitfill, J. M., & Dennison, R. (1993). Encoding and external-storage effects on writing processes. *Journal of Educational Psychology*, *85*, 267–267.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Erlbaum.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*, 564-577.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome).* New York: Academic Press.

Britton, J. (1982). Shaping at the point of utterance. In G. M. Pradl (Ed.), *Prospect and Retrospect: Selected Essays.* (pp. 139-145). Montclair, NJ: Boynton/Cook Publishers.

Bromme, R. & Stahl, E. (2005). Is a hypertext a book or a space? The impact of different introductory metaphors on hypertext construction. *Computers & Education, 44* (2), 115-133.

Connor-Greene, P. A. (2000). Making connections: Evaluating the effectiveness of journal writing in enhancing student learning. *Teaching of Psychology, 27*, 44-46.

Fulwiler, T., & Young, A. (Eds.). (1982). *Language connections. Writing and reading across the curriculum.* Urbana, IL: National Council of Tearchers of English.

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science, 21*(1), 45-71.

Graham, S. (2006). Writing. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology.* (pp. 257-278). Mahwah, NJ: Erlbaum.

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist*, *35*(1), 3–12.

Harris, K. R., & Graham, S. (1992). *Helping young writers master the craft: Strategy instruction and self regulation in the writing process.* Cambridge, MA: Brookline Books.

Hübner, S., Nückles, M., & Renkl, A. (2006). Prompting cognitive and metacognitive processing in writing-to-learn enhances learning outcomes. In R. Sun, N. Miyake & C. Schunn (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 357-362). Mahwah: Erlbaum.

Kirkpatrick, L. C., & Klein, P. D. (2009). Planning text structure as a way to improve students' writing from sources in the compare–contrast genre. *Learning and Instruction, 19*(4), 309-321.

Klein, P. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, *11*(3), 203-270.

Langer, J. A., & Applebee, A. N. (1987). *How writing shapes thinking: A study of teaching and learning.* Urbana, IL: National Council of Teachers of English.

McCrindle, A. & Christensen, C. (1995). The impact of learning journals on metacognitive and cognitive processes and learning performances. *Learning and Instruction, 5*, 167-185.

Nückles, M., Schwonke, R., Berthold, K., & Renkl, A. (2004). The use of public learning diaries in blended learning. *Journal of Educational Media, 29*, 49-66.

Rivard, L. P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching, 31*(9), 969–984.

Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67–80). New York: Guilford.

# Teacher Learning about Teacher-Parent Engagement: Shifting Narratives and a Proposed Trajectory

Corey Drake, Iowa State University, E115 Lagomarcino Hall, Ames, IA, 50011, cdrake@iastate.edu
Angela Calabrese Barton, Michigan State University, Erickson Hall, East Lansing, MI, 48824, acb@msu.edu

**Abstract:** In this paper, we examine teacher learning about teacher-parent engagement in mathematics and science education. Using data from three different schools participating in TE-PAC$^2$ (Teacher-Parent Collaborative Communities), we find evidence of teacher learning in the stories teachers tell about the meanings of – and possibilities for - teacher-parent engagement in the context of broader cultural narratives about the teaching and learning of mathematics and science in their schools and districts. In particular, teacher learning was reflected in *changes* in teacher narratives over time and in the new understandings of parent resources incorporated into these shifting narratives. We conclude with a proposed trajectory of teacher learning about teacher-parent engagement.

## Introduction

National policies (e.g., No Child Left Behind) emphasize parental involvement in schooling, but little research has been done to establish *how* teachers might engage with parents in content-based instruction, particularly in ways that acknowledge the diversity of backgrounds and resources parents bring to the classroom. This is the issue we investigated through our work in three schools with teachers and parents engaging together in science and mathematics education. As part of this effort, our research questions included:

> How and what do teachers learn about teacher-parent engagement in support of mathematics and science instruction when engaged with parents in the collaborative study of science and mathematics education reform policies, documents, and curriculum materials? How is teacher learning reflected in teachers' narratives about parent engagement in their particular schools and districts?

We investigated this question through the design and implementation of a teacher-parent collaborative experience, TE-PAC$^2$ (Teacher-Parent Collaborative Communities). At each of the three schools, this experience took on a different structure and focus, but common elements included multiple sessions of teachers and parents (along with a researcher/facilitator) working together to discuss goals for mathematics and/or science education, investigate mathematics and/or science curriculum, and jointly planning classroom or school events that engaged parents, teachers, and children together in learning and teaching mathematics and/or science.

## Theoretical Framing

In our own work, we, like others, have observed that while the rhetoric of reform—especially from multicultural perspectives—suggests that parents should be closely involved in the reform process, the reality is that parents and teachers alike struggle with how best to bring parents into reform-based instructional practices (Peressini, 1998; Remillard & Jackson, 2006). We believe that if teachers are to be successful in engaging parents in support of student learning, then it must happen within the framework of—not peripheral to—content-based instruction. In other words, teacher-parent engagement must occur not only in the context of helping with bus lines, lunch or recess duty, and field trips, but also around issues central to the teaching and learning of mathematics and science.

Our view of teacher learning is grounded in sociocultural perspectives on learning, where learning is framed as changing participation within communities of practice (Lave & Wenger, 1991). We concur with Peressini, Borko, Romagnano, Knuth and Willis (2004) that "teachers' knowledge and beliefs interact with historical, social and political contexts to create the situations in which learning to teach occurs" (p. 68). We assert that these interactions can be particularly powerful when parents are co-participants with teachers, especially when teachers and parents reflect different cultural and social backgrounds. This view of teacher learning re-situates how we think about teacher–parent engagement and how teachers learn by engaging parents. In contrast to traditional understandings of parent involvement, which include parents volunteering in a range of school-defined ways, we define engagement as an interactive process in which teachers and parents draw on multiple experiences and resources to define their interactions with one another. In prior research, we developed a model of parental engagement in urban elementary schools, specifically in relation to "what" parents engage in and "how" they do so

(Calabrese Barton et al., 2004). Parents' effectiveness in schools is related to where engagement in the curriculum occurs, as well as to the resources parents activate to support their children's schooling. This understanding of parental engagement differs from traditional models because it highlights the importance of space and capital as mediators of parental engagement and positions parents as authors and agents within schools. This last distinction is particularly important in low-income urban schools, where parents tend to feel more alienated and with little power (Harvard Family Research Project, 2002).

One area of capital exchange that has proven critical is the exchange of resources and understandings related to subject matter knowledge, pedagogy, and goals. Studies of learning in mathematics professional development indicate that teachers increased their communication skills and content understanding when engaged in content learning activities with parents (Civil & Quintos, 2002). Likewise, parents who participated in their children's elementary classrooms reported that talking about teaching, learning, and personal experiences, listening to different perspectives and opinions, and teaching mathematics to other parents were rewarding aspects of working in schools (e.g., Jackson & Remillard, 2005).

Lee and Bowen (2006) suggest that the link between parental involvement and increases in student achievement might be explained by the capital gained by parents through school involvement. However, little work has explored gains in *teachers*' capital that might also contribute to this link, especially when parental engagement is deeply situated in instructional content. It makes sense that we extend current research to include not just the nature of resources gained by parents but also those resources gained by teachers, and how the deployment of these resources frames their changing participation with parents. Thus, an integral part of our investigation is understanding teacher learning with respect to the resources that parents and families offer for a particular content area, the potential spaces within mathematics and science curricula and curriculum materials for accessing and building on these resources, and the relationship between these resources and school-based knowledge and practice.

## Data Sources and Methods

### Contexts

Our work has taken place across three distinct schools and districts, all of which are faced with intense pressures for testing in mathematics and literacy, and all of which have significant percentages of students living in poverty. The overall study took place over two years, though we spent one year in each school.

Palmer School is a fairly small school in Granite School District, with about 420 students, K-6. The school had recently taken in students from a neighboring school that closed due to lower enrollments and test scores. The principal of Palmer is highly active in recruiting novel programs to her school, such as linking her school with a fine arts center in the city, allowing the students and teachers multiple opportunities to integrate the arts across instruction. Palmer Elementary School is located in a mid-sized US city confronted with issues of changing demographics, limited resources, and testing mandates. The school district is an urban school system with approximately 20,000 students, of whom 78% are African American, 19% white and 3% Latino/a. 70% of the students are eligible to receive free lunch.

Washington Elementary, in the River City School District, is a mid-sized elementary school with three or four classrooms per grade (K-5) and approximately 550 students total. Washington Elementary was created through the merger of two schools two years prior to this study and is housed in a modern building built at that time. The building also houses a large Boys' and Girls' Club as well as a thrift store. The two schools that were merged to create Washington School were among the district's lowest achieving in terms of standardized test scores. Part of the Washington's vision is to engage parents, families, and the community in their children's education. To support this vision, Washington Elementary has a dedicated space in the school for parents and also offers several adult education opportunities. However, there is little parent involvement in classrooms or in the content of children's schooling. The student population at Washington is racially and ethnically diverse, with significant numbers of African-American, Latino/a, and Asian students, as well as students from a variety of European and African countries.

Beacon Academy is a free k-8 public charter school in Middleton. It sits on the west side of town, on the edge of two transitional neighborhoods. The school focuses overtly in its mission statement on "character" and "special" education, marketing itself to families whose children have not fared well in the regular school district. While the city of Middleton serves as home to 22% African American, 10% Hispanic, 3% Native American, and the remaining 64% White, Asian, or Pacific Islander and 23.2% children in poverty, Beacon Academy serves >95% African American or Hispanic students, and 100% in poverty. The school promotes itself on brochures and its website as a high quality school that teaches all core subjects and also offers "instruction in computers and technology, art, music, drama and Spanish.". However, due to low test scores, the school administration mandated in mid-April that no science or social studies be taught, causing some tension with our own project. Finally, Beacon

Academy's administrative staff is proud of its level of parental involvement. They hold an annual parental involvement celebration day, and the very first bulletin board one sees upon entering the school is one that welcomes parents and contains information and notices to help them navigate the school. Parents are employed on the staff to work in the office and to help out in the lunchroom. However, most parents have not been involved in the academic side of the school.

## Data Generation

We have utilized a narrative-driven ethnographic approach to investigating how learning with and about curricular materials, reform-driven documents, and academic content in collaboration with parents impacts the ways in which teachers engage parents in the design and enactment of school-based instructional practices. We utilized a range of primary methods in crafting our case studies (See Table 1).

Table 1: Overview of data generation strategies across sites

| | **Palmer** | **Washington** | **Beacon** |
|---|---|---|---|
| Teacher Interviews | 3 teachers, each interviewed twice (student artifact think aloud, interview on T-P engagement) | 5 teachers interviewed twice (beginning and end of experience), 1 interviewed once (end of TE-PAC$^2$ experience) | 4 teachers interviewed once (end of experience) |
| Parent Interviews | 4 parents, each interviewed once on T-P engagement | 2 parents, each interviewed once (Several other parents participated in TE-PAC$^2$ activities, but were not interviewed) | 2 parents interviewed once (end of experience) |
| TE-PAC$^2$ sessions | Audio and fieldnotes | Fieldnotes | Fieldnotes |
| Classroom observations | 3 teachers, science classrooms observed 6 times each, 30-60 minutes per observation, fieldnotes | Observations and field notes for Family Math Night | Observations and Fieldnotes for 7 science lessons (in science, computer, and ELA classrooms) |
| Teacher and Parent-related Artifact Collection | Lesson plans; home extension lessons; student life cycle stories | Planning sheets from TE-PAC$^2$ workshops, activities created for Family Night | Lesson plans<br>Extension activities<br>Pictures of materials for walls |
| Student Interviews and Artifacts | Student lifecycle stories<br>Fieldnotes on student participation | Artifacts<br>Fieldnotes related to student participation | 6 student interviews<br>Student storyboards<br>Student movies<br>Fieldnotes on student participation |

## Data Analysis

In analyzing and discussing these data, we utilized the Ecologies of Parental Engagement (EPE) framework we developed in a previous study (Calabrese Barton et al., 2004) and focused on the importance of space and capital. In particular, to make sense of teacher learning, we examined how participation within certain communities frames issue of space and capital. Given our work with parents in low-income urban schools, we were particularly concerned with how parents are often positioned within and against the communities of practices that make up urban schooling. This led us to a key analytic idea, which emerges from cultural-historical and hybridity theories - namely, that acts of creating hybrid spaces, discourses and identities are always political and of the highest risk for those whose knowledge, discourse, and identities are positioned as lesser. Individuals in any given community of practice engage in acts of hybridity all of the time when confronted with differences. They draw upon multiple resources or funds to make sense of the world. Yet, being "in-between" several different funds of knowledge and Discourses can be either productive or constraining, and even marginalizing, depending upon how they are recognized by those in power (Moje et al., 2004).

Case study data analysis involved multiple stages and levels of coding, all of which were based on Strauss & Corbin's (1998) procedures for open coding and method of constant comparison. In analyzing case study data, we have begun to develop coding schemes on those aspects of teaching which seem to be particularly relevant to teacher-parent engagement in urban settings, including the kinds of curricular spaces teachers create for parental engagement, the forms of capital teachers activate in support of engaging parents in content-based instruction, the roles/identities teachers take on in relation to parents, and the ways in which scientific and mathematical ways of thinking are linked to families' interests, experiences, and everyday understandings of the world (Perez Carreon, Drake, & Calabrese Barton, 2005).

In this paper, we focus in particular on the stories and counter-stories teachers tell about parent engagement in their schools. We claim that changes in these stories over time provide evidence of teacher learning related to teacher-parent engagement in mathematics and science content – including the goals, purposes, and practices that support this kind of content-related engagement. We define teacher stories, or narratives, as socially shared accounts of human intention and action (Bruner, 1990). For stories to have meaning in teaching and educational research, they must report more than a listing of dates, places and events. They must also speak to the meaning of those events in an individual's life. In the tradition developed and expanded by Clandinin and Connelly (2000) and others, we use teacher stories to develop rich, contextual, accounts of their learning. Drawing from place-based sociocultural theory, we argue that personal stories, such as teacher stories, always exist in interaction with narratives at other levels, including social and cultural (Tzou, Scalone & Bell, in press). As teachers, for example, interact with each other they do so in spaces that are shaped by cultural narratives and resources available in those places, such as a school climate deeply shaped by NCLB, curricular policies, and other boundaries and supports related to the teaching and learning of mathematics and science.

## Findings

Our findings are grounded in teacher stories and teacher practices with respect to mathematics and science instruction and parent engagement. These stories consist of the narratives that teachers craft about their teaching alongside the artifacts from practice they use as evidence for their claims. We begin with cases from two of the participating schools. (Due to space limitations, we do not present a detailed case from the third school, Beacon Academy, here, though we incorporate findings from that case in the discussion section below.) These cases provide a sense of the kinds of activities involved in the TE-PAC$^2$ experience, as well as the supports and constraints available in the different school and district contexts for this kind of work. We then describe findings related to teacher narratives – and changes in those narratives – across the three sites.

### Case 1: Palmer School

This group began meeting in February and involved three 3rd and 4th grade teachers who attended all six of our after school study group sessions, except one week when one of the teachers was absent due to a technology conference. The teachers were supported by their principal who would often stop in for a brief moment with a big smile and ask how everyone was going. They were also supported by the district curriculum coordinator for science who came to the fifth of our six sessions to ask the teachers and parents to discuss their work with her. Four parents attended fairly regularly, with one missing occasionally, and one additional parent showing up at science class because she could not make the after school sessions. One of the parents was a grandmother who had extensive parenting responsibilities for her grandchildren. Sometimes her granddaughters would come to our sessions but they mainly played with each other and tended to "ignore" what the adults were doing.

The structure of the meetings was to first to talk about "science experiences" of the week, then to "do science" together and then to discuss ways to increase parent-teacher engagement using two strategies that we proposed: Lesson Extension Ideas and Lesson Integration Ideas. Playing with science often involved taking one aspect of the mandated curriculum and doing the activities together, often in ways not prescribed in the text. The early work of the group tended to focus more on designing and implementing "extension lessons for the home" and then shifted to lessons that integrated activity in the home with activity in the classroom. The focus of these latter discussions sat within a tension of "what would parents know about science" and "what will kids do in the classroom who don't do the home component." These discussions tended to focus on the deficiencies of some of the parents within the school community, but the group decided to move forward with the belief that these activities might "bring parents in."

The major barrier that the participants seemed to face related to instructional mandates that involved science. The Granite City School District, like many around the nation, had been undergoing a series of reforms meant to help boost test scores. A recent reform initiated by the superintendent's office had been the discontinuation of any interdisciplinary teaching to ensure that each subject area was taught for the required amount of time each day. All three participating teachers were vocally opposed to the new mandate and felt "harassed" by the district official who roamed the building making unannounced visits to "check up" on instruction. This topic came up each week as teachers met to explore life cycles together – and, through these discussions, teachers and parents were able to expand their knowledge related to one another's perspective and to the broader policy and instructional contexts of school science.

What seemed particularly salient was *how* discussions of these new mandates emerged within the group and the barriers these mandates imposed for teachers and parents in their efforts to support student science learning. In particular, we noted that the teachers, from our very first TE-PAC$^2$ session onward, spoke critically of the new regulations for scripted "page a day" curricula and on the emphasis on basic skills or what the teachers referred to as "kill and drill." We also noted that the teachers contrasted these mandates with their desire to use more "hands-on" and "inquiry-based" learning in support of student motivation to learn science and making science more authentic. As Beth reminded us, students need a reason to want to understand science. Parents were less critical of the new mandates primarily because they seemed to not be as aware of the specific mandates. However, in support of the teachers' desires to use more hands-on learning, the three parents harshly criticized the culture of accountability and test scores as making it more abstract and difficult for them to understand what their students were learning and how they might better support them in learning.

## Case 2: Washington School

This group began meeting in November, 2008. Five teachers (1$^{st}$, 2$^{nd}$, 2 4$^{th}$, and 5$^{th}$ grades) showed up very regularly for all meetings, except for occasional family/child care conflicts. They had significant support from one of the Title I math teachers, who serves as a mathematics leader within the school. He worked a second job, so typically could not come to the meetings. Two out of the first three meetings were cancelled because of weather, so the facilitator and teachers decided to re-group in January.

Each meeting was attended by 2-6 parents, often with their children. The structure of the meetings was to first "do math" together (teachers, parents, and children) and then to discuss ways to increase parent-teacher engagement. Keeping in mind the "playing with science" theme from the Palmer case, most of the math provided by the facilitator for participants to do together was in the form of math activities that are not traditionally done in school – puzzles, games, tangrams, etc. There were a number of positive results and stories that emerged from this aspect of the meetings. For instance, one father was very good at a particular kind of spatial puzzle and, after observing his success with these puzzles, both the teacher and his own children began to position him as having increased expertise in mathematics. Children from the 2$^{nd}$-grade class began taking the puzzles back and forth between home and school and other children in the 2$^{nd}$-grade classroom (whose parents were not involved in the initial teacher-parent meetings) began asking for copies to take home. One family, in which the parents spoke only Spanish, worked back and forth with their child and one another to complete a Magic Square, with the child and one of the facilitators translating the "rules" of the activity while the parents provided the mathematical knowledge.

The teacher-parent-child groups generated several interesting and ambitious ideas for increasing teacher-parent engagement, including developing a school store to work on money concepts and a school garden to draw on parents' expertise in carpentry and gardening with a focus on the mathematics involved in these activities. There were frequent discussions within the groups about designing artifacts that could move between home and school in order to increase parent engagement and connections between home and school. There was considerable concern about sending things home that would not be used and ideas were generated for addressing this concern. Out of these discussions, the plans for the Family Math Nights were developed. Ultimately, the group enlisted the Title I teacher's help in putting together two Math Nights – one for 1$^{st}$/2$^{nd}$ grade and one for 4$^{th}$/5$^{th}$ grade (reflecting the grade levels of the participating teachers and parents). They, along with the Title I teacher, designed the Nights to have parents and children rotate through 4 classrooms – each with a laminated game or activity that the parents and children would play in the classroom and could then take home with them (along with the necessary dice, counters, etc.). Ultimately, these Nights were very well-attended (particularly the 1$^{st}$/2$^{nd}$-grade night) by more than 200 children and family members and teachers were very positive about this outcome.

As researchers and facilitators, this case leaves us with several conjectures and questions. In particular, Family Math Nights were not the initially-intended outcome of the project and, in that sense, the outcome is a disappointment. On the other hand, having observed the Nights and the preparation that went into them and having interviewed the teachers afterwards, the Family Nights seem to have served several positive functions related to teacher-parent engagement. Specifically, the teachers had clear ownership of this project and they continued to hold Family Math Nights during the next school year without the formal support of the TE-PAC$^2$ project.

## Teacher Stories about Teacher-Parent Engagement and Content

Initially, teachers' narratives about teacher-parent engagement were grounded in or mirrored the broader and more dominant cultural narratives of their schools. Teacher narratives, for example, focused on what it meant to teach mathematics or science in their particular school (and district) to their particular students, and what forms of parental engagement were appropriate or necessary. These cultural narratives organized teachers' discourse and practices related to what students (and parents) in their schools could accomplish, what good teaching was for their school's population, and who had the power and the right to make instructional decisions.

For example, at Washington School, teachers and parents all initially shared narratives of a general lack of parental involvement in the school, despite the school's clear mission and formal policies to engage parents and families. Some of the group's more ambitious ideas for teacher-parent engagement were rejected because of a conviction that parents at "this school" would not participate. Related to this narrative was an on-going discussion that the parents who "need" to be here (i.e., the ones whose children are struggling) were the not the ones who were participating.

This narrative at Washington School was situated in a broader cultural narrative related to the district's somewhat unique approach to mathematics teaching and curriculum. Although there was a set of standards and objectives for the district, there was no district-adopted textbook or set of curriculum materials. Instead, there was a strong and consistent narrative across the district related to teacher control over instructional and curricular choices in mathematics. This narrative was told in contrast to the district approach to literacy – a subject in which a scripted curriculum series had recently been adopted. Having an artifact in the form of common curriculum materials or even common activities/practices across classrooms might have helped focus the TE-PAC$^2$ group on examining these artifacts in order to identify "spaces" for teacher-parent engagement. The set of standards and objectives used in the district was somewhat too broad and abstract to serve this purpose.

As we can see with the teachers at Washington School, it was within these broader cultural narratives that teachers revealed how much their views of teaching and subsequent parental engagement in that teaching were framed by No Child Left Behind legislation and its extant policies and practices for mathematics and science teaching in their schools and districts. A school's status as failing or near failing and the supports and constraints this status imposed upon teachers were central to teacher narratives.

We noticed similar constraints at Beacon School. Interestingly, even when teachers had expressed views of learning that ran counter to cultural norms or formed fairly robust relationships with parents, who were highly active in the school, their more personal ideas about how teachers and parents might collaborate on instructional decisions were trumped by their perceptions of acceptable institutional practices and needs. For example, while the lead fifth grade teacher at Beacon held an expansive view of learning, which included views on how she might integrate youth interest in social networking technologies, she also spoke openly about not being able to trump the institutional narrative at her school due to the students' low test scores and school expectations for what classroom teaching ought to look like. At the same time, she had a fairly robust relationship with the one of the parents involved in TE-PAC$^2$, and yet while both she and the parent had ideas for ways to incorporate local cultural knowledge into the curriculum (i.e., incorporating oral histories or folklore into a lesson on predicting weather), they deprioritized these ideas to take on more traditional tasks supported by the broader cultural narrative. The day the lesson plan involved the students interviewing the parent on her "rules of thumb" for weather, the parent ended up not coming to science class because she felt it was more important for her to help oversee the lunch period in the cafeteria. The teacher felt she could not ask the parent to leave the cafeteria because that was the parent's "real" role in the school.

## Changes in Teachers' Narratives: Counter-Stories

We found that having opportunities to study science and mathematics together (through curriculum materials and reform documents) provided a space in which teachers began to try out narratives regarding mathematics and science instruction and parental engagement that ran counter to the broader institutional and cultural narratives. These counter-narratives were deeply grounded in the instructional content - either in how the goals for teaching content were framed, what the content of the lessons looked like, or how school and home could connect content. These counter-narratives were strikingly different from the original narratives teachers told. The original narratives focused on the spaces – or lack thereof – for parent engagement in mathematics and science instruction. As part of the TE-PAC$^2$ experience, teachers began to see new roles and identities for parents. This allowed the teachers to then imagine new narratives for teaching and learning mathematics and science. These counter-stories were supported by non-routine resources made available within these communities (e.g, family stories) and/or new meanings given to traditional resources (e.g., Family Math Nights). Finally, these counter narratives were co-authored and constructed/re-constructed over time as parents and teachers took greater risks in sharing ideas and perspectives (e.g., building an understanding of and then an opposition to district policies over several weeks as parents and teachers realized they shared similar concerns).

Engaging in the study of science and mathematic curriculum and reform materials supported opportunities for teacher learning by shifting the spaces in which teachers and parents typically engage each other. Instead of interacting over individual student issues or becoming involved in the management of the classroom, studying science and mathematics curriculum and reform materials together provided opportunities for sustained dialogue on science and mathematics in everyday lives and shifted discourse from expert/novice to co-learners. In other words, these shifts in spaces altered the resources and roles available to parents and teachers allowing them to take up alternative narratives that countered the cultural norm. Teachers' views on resources for learning shifted as they began to view parents in a more collaborative way. Some of the teachers expanded their repertoire of resources to support their teaching and student learning when they began to understand that parents were more than recipients of information or experts only on their own kids.

At all of our sites, parents and teachers' collaboration eventually served as a somewhat safe space for critical discourse on school policies, and in two cases such discussions led to opportunities for teachers and parents to subvert aspects of school policy that they collectively agreed worked against children's learning. In both cases the policy subverted related to the allocation of instructional time: In one case (Palmer) ignoring the superintendent's call for no interdisciplinary teaching, and in a second case (Beacon) ignoring the school mandate to eliminate all instructional time for science. At the third site (Washington), these discussions led teachers and parents to carve out a space for parent engagement in mathematics that was notably and explicitly resistant to the dominant curricular and parent engagement policies that had been created with respect to literacy. Also at Washington, the high attendance at the Nights provides a "counter-story" to the clear belief and expectation of both teachers and parents that parents at this school do not participate in school events, particularly those related to content.

## A Cycle of Teacher Narratives

If we think about teacher learning about teacher-parent engagement as a trajectory, we conjecture teachers made important movement along the trajectory, but that movement was mediated by the social context (i.e., school level support for science, teaching morale) along with teachers' personal resources for learning (i.e., teaching identity and knowledge of content). Teachers in all three schools experienced a culture of regulation and disempowerment and/or exhaustion among teachers. In fact, this trajectory cannot be understood separate from trajectories related to teachers' understandings of what it means to teach mathematics and science in their particular school and district. In proposing a trajectory of teacher learning about teacher-parent engagement, we claim that the center of this trajectory is defined by a cycle of teacher narratives (Figure 1) in which narratives of teacher-parent engagement are initially situated in broader cultural narratives about mathematics and science education and the role of parents. However, as teachers exchange resources with parents in the spaces defined by the TE-PAC$^2$ experience, they begin to tell counter-narratives not only about teacher-parent engagement, but also about the teaching of mathematics and science more broadly. In other words, teachers' narratives shift from questions of where parents might "fit" within the traditional teaching of mathematics and science to how new understandings of parent engagement might transform the teaching and learning of mathematics and science.

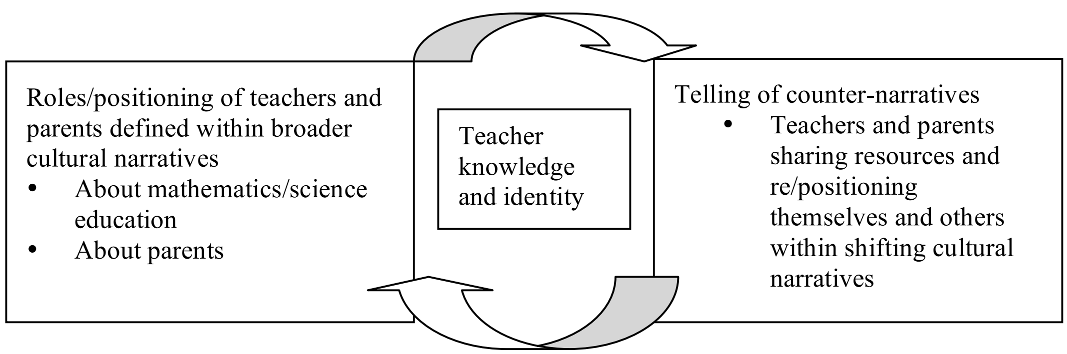



Figure 1. Cycle of Teacher Narratives.

At Beacon, we see the teachers as positioned at an early point along this trajectory. The teachers at Beacon were less familiar with the new science standards and also had no school-purchased resources for science. They tended to use the environment as a shield to prevent further engagement with parents. The initial narrative that emerged among teachers and parents in the context of *content* explorations focused on the "troubles of the school" that positioned teachers and parents differently and sometimes in opposition rather than the content of instruction.

Teachers at Washington and Palmer appeared to be positioned further along the trajectory as they were better able to take advantage of opportunities to learn with and through parental engagement and engage in the iterative cycle of narratives depicted in Figure 1. For example, at Palmer, teachers used the TE-PAC$^2$ experience as an opportunity to cultivate allies to work against policies that they felt worked against meaningful student learning. They used the TE-PAC$^2$ space to share complaints and to strategize on how to work against the negative environment. At both Washington and Palmer, teachers' relationships with parents as people with experiences that matter in science and mathematics developed dialectically with their views on the kinds of resources that support student learning. Consequently, teachers shifted classroom practice to reposition parents with more authorship in the classroom and/or curricular space. Parents helped to design extension lessons and served as content experts. They also provided feedback on their children's responses to the classroom. However, in this hybrid community, teachers still maintained overt control over when and how parents' shifting roles became public, even though many of the teachers expanded their views on what was possible in their school or classroom. Going forward, we hope to continue to understand, both theoretically and practically, the nature of this trajectory and ways to support teachers and parents in moving along this trajectory towards creating broader spaces and capital exchanges for engagement in the teaching and learning of mathematics and science.

## References

Calabrese Barton, A., Drake, C., Perez, G., St. Louis, K., & George, M. (2004). Ecologies of parental engagement in urban education. *Educational Researcher 33*(4): 3-12.

Bruner, J. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Civil, M. & Quintos, B. (2002). *Uncovering mothers' perceptions about the teaching and learning of mathematics.* Paper presented at the Annual Meeting of the American Education Research Association. New Orleans, LA.

Clandinin, D.J. & Connelly, F.M. (2000). *Narrative Inquiry: Experience and Story in Qualitative Research.* San Francisco: Jossey-Bass Publishers.

Harvard Family Research Project (2002). *Concepts and Models of Family Involvement.* Cambridge, MA: Harvard University.

Jackson, K., & Remillard, J. T. (2005). Rethinking parent involvement: African American mothers construct their roles in the mathematics education of their children. *The School Community Journal, 15*(1), 51-74.

Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation.* Cambridge: Cambridge University Press.

Lee, J-S. & Bowen, N.K. (2006). Parent involvement, cultural capital, and the achievmenet gap among elementary school children. *American Educational Research Journal 43*(2): 193-218.

Moje, E. B., Ciechanowski, K. M., Kramer, K., Ellis, L., Carrillo, R., & Collazo, T. (2004). Working toward third space in content area literacy: An examination of everyday funds of knowledge and Discourse. *Reading Research Quarterly, 39*(1), 38-70.

Peressini, D.D. (1998). The Portrayal of Parents in the School Mathematics Reform Literature: Locating the Context for Parental Involvement. *Journal for Research in Mathematics Education 29*(5), 555-582.

Peressini, D. Borko, H. Romagnano, L., Knuth, E. & Willis, C. (2004). A Conceptual Framework for Learning to Teach Secondary Mathematics: A Situative Perspective. *Educational Studies in Mathematics 56*(1): 67-96.

Perez Carreon, G., Drake, C. & Calabrese Barton, A. (2005). The importance of presence: Immigrant parents' experiences of school engagement. *American Educational Research Journal 42*(3): 465-498.

Remillard, J. & Jackson, K. (2006). Old Math, New Math: Parents' experiences with standards-based reform. *Mathematical Thinking and Learning*, 8(3), 231-259.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: grounded theory procedures and techniques*. Thousand Oaks, CA: Sage.

Tzou, C., Scalone, D. & Bell. P. (in press). The role of environmental narratives and social positioning in how place gets constructed for and by youth. *Equity & Excellence in Education.*

## Acknowledgements

This work was supported, in part, by a grant to the two authors from the Spencer Foundation. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the Spencer Foundation.

# Effects of Instructional Design Integrated With Ethnomathematics: Attitudes And Achievement

Melike Kara, Illinois State University, Cardinal Court, B21, 61761, Normal, IL, mkara@ilstu.edu
Aysenur Yontar Togrol, Bogazici University, Ortaogretim Fen ve Matematik Alanları Egitimi Bolumu, 34342 Bebek Istanbul, Turkey, yontar@boun.edu.tr

**Abstract:** Development of an instruction integrated with ethnomathematics and its effects on 7th grade students' attitudes towards mathematics& achievement levels are the main concerns of this paper. Participants were selected from the 7th grade students. Six hours of instruction consisting of transformational, reflectional and rotational symmetry, patterns and ornaments based on the mathematics used in Topkapı Palace were presented to treatment group. An attitude scale, two achievement tests were administered as pre and post tests. Repeated measures two way ANOVA was used in order to test the hypothesis.

## Introduction

Schools tend to ignore the ethnomathematical knowledge that the children themselves can bring into the classroom from their homes and communities although math is not a culture-free discipline (Zaslavsky, 1998), therefore mathematical concepts and ethnomathematical perspectives should be synthesized in the curriculum in a way which is sensitive to all cultures. Development of an instruction integrated with ethnomathematics for 7$^{th}$ grade math lessons and its effects on students' attitudes towards mathematics and achievement levels are the main concerns of this paper. Practical mathematics and scholarly mathematics are tried to be narrowed with the instruction developed in the study.

D'Ambrosio (2001) states that ethnomathematics encourages us to witness and struggle to understand how mathematics continues to be culturally adapted and used by people around the planet and throughout time. In traditional math classes there exists no connection between culture and mathematic. As a result of this practice students believe that there is no connection between them and math is "acultural" which means that a discipline without cultural significance. Researches show that achievement level in math is highly correlated with students' attitude towards math. Butty (2001) noted that students with better attitudes towards math had a significantly higher achievement scores than those with poorer attitudes toward mathematics. Schultes and Shannon (1997) found that many students gained greater appreciation for math after learning the subject matter from a cultural perspective since it made students more comfortable and confident about discussing mathematical concepts such as infinity with their peers. (Cited in Arishmendi-Pardi, 2001).

## Literature Review

All of the definitions of ethnomathematics consider the points such as the culture in which mathematics arise, mathematics that is implicit in cultural in cultural practices, the use of anthropological principals to investigate mathematical practices, relation and implications for mathematics education. The definition of D'Ambrosio, which is the mathematics practiced among identifiable cultural groups, such as national-tribal societies, labor groups, children of a certain age bracket, professional classes (D'Ambrosio, 1985; cited in 1997, p. 16) is used as the main focus of the instruction integrated with ethnomathematics. Besides Zaslavsky (1998) stated that ethnomathematics includes mathematics of people's ethnic and racial group's, various classes in the society and practices of students' own communities.

The researches on ethnomathematics attempts to supplement the existing curriculum through the cultural practices and or/and examine the outcomes of integrating ethnomathematics into classroom instruction. There are many studies that seek to supplement the curriculum and their focus is usually on indigenous cultures (e.g. Barta, Abeyta, Gould, Galindo, Matt, and Seaman, 2001; Gerdes 1988; Lipka, Hogan, Webster, Yanez, Adams, Clark, and Lacy, 2005; Morales 1993). Studies that focused on outcomes did this through an examination of cultural practices and or artifacts, students' home/everyday experience or students' cultural practices. In general related research studies reach to similar conclusions that integrating ethnomathematics into mathematics instruction has positive effect on students performances (Arishmendi Pardi, 2001; Lipka et. al, 2005; Brenner, 1998).

The literature review showed that people use the mathematical activities in their real life with cultural practices, artifacts (Morales, 1993). In everyday activities, people involve a substantial amount of mathematical applications, which are universal behaviors such as counting, measuring, designing, locating, explaining, playing and for example, when designing tools, patterns are used especially for clothing, weapons (Bishop, 1991 as cited in Barta et. al, 2001). Moreover Moses (2005) noted that students' achievement levels in mathematics increased as they learn about African culture. The grades of algebra course of students taught ethnomathematical pedagogy were higher than the students' who took without the ethnomathematical pedagogy (Arishmendi-Pardi, 2001). The connection between mathematical ideas and lived experience of individual students, were emphasized in the study of Presmeg (1998) which argues that the ethnicity of students is a resource for mathematics. Powell and Temple (2001) say that Oware which is a board game played in Africa can help children understand that humans encode their mathematical ideas in diverse cultural products, including architecture, art games, music and written texts. Gerdes (1998) argued to develop an awareness of the social and cultural bases of mathematics, which contributes both to enhance self-confidence, capacity, readiness and openness to work in a multicultural environment among future math teachers. Teachers can make students see that in their activities, there are mathematics engaged in and they know math (Gerdes 1998). Barkley and Cruz (2001) noted that Native American beadwork incorporated in daily lives shows a high degree of sophistication in terms of specific symmetrical patterns which are used to symbolize the balance and communicate people's feeling of harmony with the natural world in which they lived. Gerdes (2001) suggested that teachers may look for suitable activities from diverse cultural contexts and analyze how these activities may be integrated into their teaching to create a truly simulating and enriching environment to help all students fully develop their potentials. Lipka and his colleagues (2005) found that the results of students' mathematics performance were in favor of the instruction which is math in a cultural context and based on two case studies of a successful culturally based math project.

In the light of the studies conducted on ethnomathematics, the purpose of this study is to investigate the effects of math instruction integrated with ethnomathematics. Specifically to find out if there will be any statistically significant difference between the level of attitude towards mathematics and mathematics achievement levels of seventh grade students who receive instruction integrated with ethnomathematics and regular instruction.

## Methodology

All of the definitions of ethnomathematics consider the points such as the culture in which mathematics arise, mathematics. In this section, the experiment set-ups and procedures will be described. These include the participants, the instruments, the data collection process, and the analytical procedure.

### Participants

Fifty students in seventh grade participated in this study. Due to practical reasons, convenient nonrandom sampling technique was used. The school is a state supported school, which has two seventh grade classes. One of the classes was control group (n=30) treated with the regular instruction and the other one was treatment group (n=20), treated with the instruction integrated with ethnomathematics.

### Design

In this quasi experimental study, since the participants were selected non-randomly except the random assignment of intact groups to treatments the design may be called as non-equivalent control group design (Gay, 2003). Before the treatment, Mathematics Achievement Scale1 (MAS1) and Mathematics Attitude Scale (MATT) were administered as pretest to all control and treatment groups. After the treatment, in order to find out achievement and attitude differences MATT scale and Mathematics Achievement Scale2 (MAS2) were administered as post tests.

### Instruments

In order to determine students' attitude towards mathematics at the beginning and at the end of the study, students were given mathematics attitude scale (MATT) which was developed by Nazlıçiçek and Erktin (2002). The reliability of MATT was calculated with the Cronbach Alpha Coefficient which is .84 for the sample 234 students from 6th, 7th and 8th grades. For the content validity of the test, factorial analysis of variances was conducted and in three subscales -which are mathematics achievement level perceived by the students, the benefits of mathematics perceived by the students and interest toward math lessons- 52 percent of the variance explained. For construct validity of the scale, the correlation coefficient between students' attitudes towards mathematics and their math achievement is found as .36 which is statistically meaningful for .01 significance level. According to Nazlıçiçek and Erktin (2002) this value was consistent with the previous studies Minato and Yanese, (1984); Ethington and Wolfle, (1986); Cheung (1988); Erktin, (1993).

In order to measure students' knowledge on the prerequisite topics for the symmetry and patterns (angle, line and plane using protractor to measure angles, basic characteristics of polygons), Math Achievement Scale 1(MAS1) was developed and used as a pre-test. Mathematics Achievement Scale 2 (MAS2) was developed for the line symmetry, rotational symmetry and patterns topics and used as a posttest. Both of instruments were developed by the researchers. The questions were adapted or directly taken from nationwide Teacher Guide Book of 7th Grade (Aygün, S.Ç., Aynur, N., Çuha, S.S., Karaman, U., Özçelik, U., Ulubay, M., Ünsal, N., 2007). For content validity of the instruments two mathematics teacher with master degree judged whether they evaluate students' knowledge on the prerequisites of symmetry and patterns topics for MAS1 and all subtopics of the symmetry and patterns topics for MAS2.

## Procedure

This section includes the description of the treatments, which are instruction integrated with ethnomathematics and regular instruction in "Mathematics in Our Daily Life" chapter. Both experimental and control treatments lasted in six-lesson hours. All of the instructions were implemented by the same mathematics teacher who is one of the researchers.

Instructions were prepared for the symmetry and patterns topics of the $7^{th}$ grade math lessons. Both of them are developed according to the objectives of $7^{th}$ grade mathematics curriculum of National Education of Turkey, as stated in Teacher Guide Book of $7^{th}$ Grade (Aygün, et al., 2007).

The regular instruction was based on the activities stated in the text-book. The instruction integrated with ethnomathematics for the $7^{th}$ grade symmetry and patterns topics aimed to enrich the existing curriculum with the cultural artifacts and ornaments. This type of instruction was based on the second strand of ethnomathematics defined as the analysis of the mathematics of the traditional cultures and of indigenous people and the fourth strand which focuses the relationship of ethnomathematics with the formal education system (Vithal and Skovsmose, 1997). It was designed according to mathematics of the tilling and patterns of the rooms called as Murat III Private Room, Mother of Sultan Flat in Topkapı Palace, practiced by the carpenters and tilling experts and the reasons behind the usage of the mathematics in those rooms by integrating to the instruction stated Teacher Guide Book of $7^{th}$ Grade Chapter 5, prepared according to mathematics curriculum of National Education of Turkey. Students discoverd the use of symmetry and patterns Turkish ornaments and Topkapı Palace which has an important role in Turkish History and full of symmetric ornaments. After a PowerPoint and a video show with photographs was shown, the selected rooms in the Harem section of Topkapı Palace, were examined with a simulation named as "360 degree" which makes use to see all the room within different perspectives. The motifs and ornaments are examined one by one with their photographs in the presentation and in the worksheets. Students made activities to define pattern types in the ornaments of to answer the questions: *Does the shape have rotational symmetry? If yes, what is the minimum angle that we should rotate to have the shape in the same posisiton? Does the shape have reflectional symmetry? Does the shape have reflectional symmentry more than one axis? Is there any translation of the main motif?* Then social and architectural properties of Harem Section was explained to students. Some of the properties are as follows:

*This is the house of the Ottoman emperor where his family lives so reflects the family life of the emperor. It was forbidden to enter the foreigners to enter Harem Section. Harem section was dark and cold since there were very few windows because of its secrecy. There was a hierarchical placement in the rooms and the most important people rooms' ornaments had higher complexity that the less important ones. The ornaments do not include animated figures because of religious reasons. Geometrical figures and flowers are used in tilling, cushions and carpets. The pools and taps were to cool off in the summer, and also for hygienic reasons. Another usage was not to make others listen outside of the room. When somebody is talking, another person outside of the room cannot hear anything if the tap was opened. The basic living problems were secrecy, light and heating. To make the rooms hotter there were fireplaces. The walls were full of tilling with symmetrical figures to reflect the heat and to foreclose humidity. There were candle places in the walls and they were also full of those figures. The carpets on the ground and hang in front of doors and windows were also to avoid cold weather. In the rooms, there were three types of ornaments which are calligraphy, flower deigns and patterns with repeated polygons-tessellations. The third type was related with the instruction integrated with ethnomathematics. Their basic characteristics is having infinitely continued, symmetrical and n-sided.*

The main difference between the instruction integrated with ethnomathematics and the regular instruction is the shapes used in the activities. The regular instruction activities, as stated below consist of the shapes and examples in the geometry. On the other hand in the activities of the instruction integrated with ethnomathematics, the shapes are chosen from The Flat of Valide Sultan and Room of Murat III. The other difference is that students in the experimental treatment also learn about Turkish history and lifestyles of the people who were living in the palace in the past centuries. The worksheets which were implemented as class exercises during experimental treatment are parallel to the students' practice and activity parts stated in the Teacher Guide Book of 7th Grade Chapter 5. However the shapes were selected from the patterns and ornaments of Room of Murat III and The Flat of Valide Sultan. In addition, they were prepared in the ethnomathematical perspective.

Both of the instructions include the use of the symmetry and patterns in daily life. In regular instruction it is presented with only examples, without any explanation or reason about the usage of them. The instruction integrated with ethnomathematics on the other hand includes the use of symmetry and patterns with reasons in a cultural context.

## Data Analysis and Results

Firstly, the distribution of the a) MATT as a pretest and b) MATT as a posttest c) MAS1 and d) MAS2 scores of the students in the treatment and control group were determined. Normality tests of Kolmogorov-Smirnov and Shapiro-Wilk showed that these scores are not significantly different than the scores which have normal distributions.

Since two dependent variables are measured repeatedly on independent groups where each group is exposed to a different condition Repeated Measures Analysis of Variance (ANOVA) was conducted for control and treatment groups. In order to determine whether control and treatment groups are significantly different or not in terms of their attitudes towards mathematics (MATT test scores) and mathematics achievement levels (MAS1 scores and MAS2 scores) at the end of the treatments, Repeated Measures Two Way ANOVA were carried on the pretest and posttest scores of two groups for both of the variable. Descriptive statistics are presented in Table 1.

Table 1: Descriptive Statistics for Attitude towards Mathematics and Mathematics Achievement.

| | Control Group | | Treatment Group | | TOTAL | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| N | 30 | | 20 | | 50 | |
| **Pretest** | | | | | | |
| MATT | 41,97 | 12,694 | 55,50 | 20,433 | 47,38 | 17,382 |
| MAS1 | 66,83 | 14,130 | 66,00 | 16,746 | 66,50 | 15,069 |
| **Posttest** | | | | | | |
| MATT | 33,32 | 10,346 | 58,42 | 15,812 | 43,36 | 17,739 |
| MAS2 | 64,43 | 13,756 | 67,05 | 13,193 | 65,48 | 13,459 |

First hypothesis of the study was; *there will be statistically significant difference between the mathematics achievement levels of seventh grade students who receive instruction integrated with ethnomathematics and regular instruction* was not supported. Repeated Measures Two way ANOVA for control and treatment groups' scores on MAS scores revealed that ($F=1.812$, $p=.185$) the difference between the groups in terms of mathematics achievement was not statistically significant.

Second hypothesis of the study; *There will be statistically significant difference between the attitude levels of seventh grade students who receive instruction integrated with ethnomathematics and regular instruction* was supported by the results of the statistical analysis. Repeated Measures Two way ANOVA for control and treatment groups' scores on MATT indicated that $F=8.295$ with $p=.006$. The difference between the groups is statistically significant ($p=.006$) in favor of treatment group.

## Conclusion and Implications

The study showed that there were no significant effects of instruction integrated with ethnomathematics on mathematics achievement. This finding of the study contradicted with the report of Moses (2005) that African-American students' achievement scores in mathematics increased as they learn about African culture. Also the grades of algebra course of students taught ethnomathematical pedagogy were higher than the students' who receive instruction without the ethnomathematical pedagogy (Arishmendi-Pardi, 2001). Moreover, many researchers reported the superiority of integrating ethnomathematics into mathematics instruction on students' mathematics performance (e.g. Arishmendi Pardi, 2001; Lipka et. al, 2005; Brenner, 1998,). Another contradiction was found with Lipka and his colleagues (2005) findings that the results of students' mathematics performance were in favor of the instruction which is math in a cultural context and based on two case studies of a successful culturally based math project. So results of the current study about achievement levels may be questioned in terms of the time/length of the treatments or about the differentiation of the treatments.

Second finding of the study showed that instructional practices which was enriched by ethnomathematical examples effected students' attitudes towards mathematics positively in treatment group.
This study provides an argument for a teaching technique, instruction integrated with ethnomathematics to improve math education. The findings suggested that, integrating ethnomathematics is an effective strategy to promote students' attitudes towards mathematics. In order to enhance self-confidence, capacity, readiness and openness to work in a multicultural environment among future math teachers Gerdes (1998) argued to develop an awareness of the social and cultural bases of mathematics.

Additionally, this study provides helpful findings on integrating ethnomathematics to mathematics lessons within a mathematics learning setting. The connection between mathematical ideas and lived experience of individual students, were emphasized in the study of Presmeg (1998) which argues that the ethnicity of students is a resource for mathematics. As Gerdes (2001) suggested teachers may look for suitable activities to create a truly simulating and enriching environment to help all students fully develop their potentials.

The ornaments in Topkapı Palace prompted students to look at the colors, shapes, and patterns in the culture and reinforced the idea that ornaments can be a means of expressing one's cultural heritage, living style, psychology and status. The exploration of different ornaments, ethnomathematical practices, tilling and patterns allows students to begin to see how cultures express themselves and realize how the mathematics that they use in school is connected with things that they see in the world. This presentation offers a few ideas for using ornaments to help students explore various cultures and mathematical concepts. These activities can be modified for use with students in other grades. As students arrange and learn the characteristics of the shapes to form the tilling, they can explore the importance of the number of sides, angles, and so forth. Teachers might also modify the lessons by searching other ethnomathematical practices.

Finally in this paper, although the influence of instruction integrated with ethnomathematics was not statistically significant, in general a potential trend showing the higher mathematics performance was observed in other researches investigating the effects of instruction integrated with ethnomathematics. This is an interesting trend deserving further research.

## References


Arishmendi-Pardi, E. J. (2001). Comparison of the final grades of students in intermediate algebra taught with and without and ethnomathematical pedagogy. A presentation to the Center of Diversity in Teaching and Learning in Higher Education, Miami, FL. Retrieved August 26, 2007 from http://www.rpi.edu/~eglash/isgem.dir/texts.dir/ejap.htm.

Aygün, S.Ç., Aynur, N., Çuha, S.S., Karaman, U., Özçelik, U., Ulubay, M., Ünsal, N. (2007). *Öğretmen Kılavuz Kitabı Matematik 7.*İstanbul: Feza Gazetecelik A.Ş. MEB Devlet Kitapları.

Barta, J., Abeyta, A., Gould, D., Galindo, E., Matt, G., Seaman, D. (2001). The mathematical ecology of the Shoshoni and implications for elementary mathematics education and the young learner. *Journal of American Indian Education*, 40(2), 1-27.

Barkley, C. A. & Cruz, S. (2001). Geometry through beadwork designs. *Teaching Children Mathematics*, 7(6), 362-367.

Brenner, M. E. (1998). Adding cognition to the formula for culturally relevant instruction in mathematics. *Anthropology and Education Quarterly 29(2), 214-244.*

Butty, J.L. (2001). Teacher Instruction, students' attitudes and mathematics performance among 10th and 12th grade Black and Hispanic students. *The Journal of Negro Education:* Winter 2001, 70.

D'Ambrosio, U. (1997). Ethnomathematics and its place in the history of pedagogy of mathematics. In A.B. Powell & M. Frankenstein (Eds.), *Ethnomathematics: challenging Eurocentrism in mathematics education* (pp. 13-24). Albany, NY: State University of New York Press.

D'Ambrosio, U. (2001). What is ethnomathematics, and how can it help children in schools? *Teaching Children Mathematics*. Reston: 7(6) 308.

Gay, L. R. & Airasian P. & Miles G. (2003). *Educational Research.* Upper Saddle River: Merrill/Prentice Hall.

Gerdes, P. (1998). On culture and mathematics teacher education. *Journal of Mathematics Teacher Education,* 1: 33-53.

Gerdes, P. (2001). Exploring the game of "Julirde": A mathematical-educational game played by Fulbe children in Cameroon. *Teaching Children Mathematics.* 7(6), 321-327.

Lipka, J.& Hogan, M. &Webster, J.B. & Yanez, E. & Adams, B. & Clark, S. & Lacy, D. (2005). Math in a cultural context: Two case studies of a successful culturally based math project. *Anthropology and Education Quarterly,* 36; 4 pg 367-385.

Ma, X., & Kishor, N. (1997). Attitude toward self, social factors, and achievement in mathematics: A meta-analytic review. Educational Psychology Review, 9(2), 89-120.

Ma, X., Xu, J. (2004). Determining the Causal Ordering between Attitude toward Mathematics and Achievement in Mathematics. *American Journal of Education.* (v110 n3.)

Matematik6_8_25082006, Mathematics Curriculum of Primary Schools in Turkey, Retrieved January 15, 2008 from http://ttkb.meb.gov.tr//indir/ogretmen/programlar/program_son/matematik6_8_25082006.zip

Morales, L. (1993, December). Mayan Geometry. *ISGEM Newsletter, 9(1).*
Retrieved August 26, 2007 from
http://web.nmsu.edu/~pscott/isgem91.htm

Moses-Snipes, P.R. (2005). The affect of African culture on African American students' achievement on selected geometry topics in the elementary mathematics classroom. Negro *Educational Review,* 56, 2(3) pg 147.

Nazlıçiçek, N & Erktin, E. (2002) *İlköğretim Matematik öğretmenleri için kısaltılmış matematik tutum ölçeği.* V.Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, Ankara. Retrieved March 13, 2007; from http://www.fedu.metu.edu.tr/UFBMEK-5/b_kitabi/PDF/Matematik/Poster/t194.pdf

Powell, A. & Temple O. (2001). Seeding Ethnomathematics with Oware: Sankofa. *Teaching Children Mathematics.* 7(6), 369-375.

Presmeg, N.C. (1998). Ethnomathematics in Teacher Education. *Journal of Mathematics Teacher Education,* 1, 317-319.

Vithal, R. & Skovsmose, O. (1997). The end of Innocence: A critique of ethnomathematics. *Educational Studies in Mathematics.*34, 131-157.

Zaslavsky, C. (1998). Ethnomathematics and multicultural mathematics education. *Teaching Children Mathematics*, 4(9), 502-503.

# The impact of web-based collaborative inquiry for science learning in secondary education

Annelies Raes, Tammy Schellens, Bram De Wever, Department of Educational Studies, Ghent University,
Henri Dunantlaan 2, 9000 Ghent,
Email: annelies.raes@ugent.be, tammy.schellens@ugent.be, bram.dewever@ugent.be

**Abstract:** Current educational practice in higher education shows a growing use of CSCL-environments. In secondary education, however, we only see the first signs of transformation in the age of technology and the Internet, and many of these signs are particularly at the administrative level rather than in the classroom. Nevertheless, research, as well as national standards, support collaborative learning and the integration of ICT as an answer to the decreased interest and motivation in science learning and the growing importance that is attached to inquiry skills. This research project deals with the use of web-based collaborative inquiry as a promising approach for secondary science education. In particular, this study investigated the impact of the implementation of a Web-based Inquiry Science Environment (WISE) on students' content knowledge, their inquiry skills and their attitude and engagement towards science. An empirical study in 19 secondary science classes was conducted and 375 students were involved. Additionally, this study focused on gender differences and highlighted the transformation of the teachers' role in web-based teaching. The present study demonstrates the effectiveness of this innovative instructional approach in the attempt of making science accessible and interesting to all and to rectify the gender imbalance in science education.

## Introduction

The latest Eurobarometer on "Young People and Science" (2008), investigating the public's opinion, revealed that young Europeans (aged between 15 and 25) have a positive view about science and technology. However, when presented with several choices of scientific study, only a minority of these young people said they were considering them. These findings are in line with the first results of PISA 2006. The PISA results reveal that Flanders belongs to the group of OECD countries which achieved very high results for scientific literacy. On the other hand, in comparison with the 15-year-old students in the average OECD country, fewer Flemish students reported that they are motivated to learn science, and only an absolute minority thought that they will work with science later on (De Meyer, 2008). The number of students who consider taking up studies and careers in science is at a low level. This is especially true for female students. Recent research findings emanating from a range of countries demonstrate that gender equity in science education is still a cause for concern. One of the reasons for this gender differentiation in science interest is the perception shared by many girls that only boys or men are good in science and technology (Sedeno, Balmaseda, Suarez, & Dauder, 2008). Nevertheless, different research findings demonstrate that during elementary education girls' scientific achievement is equal to boys' achievement. However, in secondary education, boys significantly outperform girls with respect to science (TIMMS, 2003; PISA, 2006). The number of girls that graduate in a scientific, mathematic or technical master is lower than the number of boys. Moreover, according to the Flemish board for scientific policy (2008), this number is still decreasing.

Concerns are raised about this lack of interest in science, and the resulting reduction in the numbers of young people who opt to study science subjects at an advanced level. These concerns seem justified, especially in an industrialized society. In addition to potential economic and labor market consequences of this disinterest, there are also social and cultural consequences, since science represents an important part of the culture in which an understanding of scientific phenomena is required to participate fully in many topical debates (Woodgate & Stanton Fraser, 2007). One of the reasons for young people's lack of interest in science is that much of what goes on in science classrooms is not particularly attractive to students (Stark and Gray, 1999; Flemish governmental enquiry, 2005).

Next to that, National standards stress the growing importance that is attached to inquiry skills. More particularly, enhancing student's understanding of science concepts and process skills rather than merely teaching the lower textual-level scientific knowledge is a major goal for science educators (Galili, 1996).

These findings stress the review, rethink, and reform of science education to make science accessible and interesting to all and to rectify the gender imbalance in science education. In this respect, we investigated the impact of web-based collaborative inquiry as an innovative instructional approach for science learning in Flemish secondary education.

## Theoretical and Empirical Framework

Inquiry-oriented science instruction has been characterized in a variety of ways over the years (Collins, 1986; DeBoer, 1991; Rakow, 1986) and promoted from a variety of perspectives. Inquiry based learning is a student-centered, active learning approach which stimulates students to get involved in a social, active, engaged, and constructive learning process, as opposed to more traditional approaches which tend to emphasize the memorizing of factual information. According to constructivist models, learning is the result of ongoing changes in our mental frameworks as we attempt to make meaning out of our experiences (Osborne & Freyberg, 1985).

Inquiry-based programs have been found to generally enhance student performance related to scientific literacy and understanding of science processes (Lindberg, 1990), conceptual understanding (Slotta & Linn, 2009) and positive attitudes toward science (Kyle et al., 1985; Rakow, 1986). Moreover, research shows clearly that by reflecting, applying ideas, and collaborating with peers, students develop a sense of the relevance of science (Bransford, Brown, & Cocking, 1999). Research confirms the commonly held view that collaborative learning motivates students (Gillies, 2003) and there is evidence for the fact that students develop a positive attitude towards the course in which collaborative learning is applied (Stevens & Slavin, 1995; Nichols & Miller, 1994; Nichols, 1996; Springer, Stanne & Donovan, 1999).

With the growing importance of the World Wide Web, this information resource can serve as a means for collaborative inquiry that opens the boundaries of the classroom and creates the possibility for students to pursue questions of personal interest (Wallace, Kupperman, Krajcik, & Soloway, 2000). Research on Computer Supported Collaborative Learning (CSCL) has contributed to the claim that collaborative activity among students can effectively be supported with computer technology (Lou, Abrami, & d'Apollonia, 2001). In this respect, web-based collaborative inquiry is a promising approach. National standards and policy advisors confirm this and call for inquiry learning and the integration of technology into science classes.

In higher education, we are now witnessing a slow but steady growth in the use of CSCL, where student activities and student-teacher exchanges are coordinated through online environments. In primary and secondary education, however, we only see the first signs of transformation in the age of the Internet, and many of these signs are particularly at the administrative level rather than in the classroom (Cox, Abbott, Webb, Blakeley, Beauchamp, & Rhodes, 2004; Tondeur, van Braak, & Valcke, 2007). Apparently, teachers prefer conservative, rather than revolutionary applications of technology. Teachers remain predominantly focused on lectures and textbooks, using the Internet primarily as a supplemental resource for Web searches or multimedia materials (Slotta, 2004). Nevertheless, in research institutions technology-based environments for (collaborative) inquiry learning have been developed, such as WISE (Slotta, 2004) , Co-LAB (van Joolingen, de Jong, Lazonder, Savelsbergh, & Manlove, 2005) or BGuILE (Reiser, Tabak, Sandoval, Smith, Steinmuller, & Leone, 2001) as an arguably, more interesting and motivational approach to secondary science education.

Although different technology-based environments for collaborative inquiry learning have been developed, little large-scale research has been conducted with regard to the benefits of this approach. Through an intervention study we investigated the impact of a web-based collaborative inquiry project on content knowledge, inquiry skills and attitude and engagement towards science.

## Research questions

Four major research questions drove this study:

1. What is the impact of the web-based collaborative inquiry project on students' understanding of scientific phenomena?
2. What is the impact of web-based collaborative inquiry on students' inquiry skills, particularly hypothesis generation?
3. What is the impact of web-based collaborative inquiry on students' attitude and engagement towards science?
4. Is there a differential impact for gender?

Additionally, this study wanted to investigate web-based collaborative inquiry from the teachers' point of view with the aim of exploring how the changed classroom activity created by being online and involved in an inquiry project affected the work of teaching.

## Research design and Method

### Participants

The participants in this study were 375 students from 19 secondary school classes (grade 9, 10, and 11, the average age of these students was 16 years). The ratio of males to females in the participants was 54% girls to

46% boys. The classes were selected from 15 Flemish secondary schools and a group of 17 science teachers were involved in the research project. Teacher participation in the intervention was voluntary and teachers agreed to dedicate 4 lessons of 50 minutes to implement the CSCL-project.

## Procedure and design

Forty-six master students Educational Sciences were involved in this study to support the implementation of the project and to conduct the questionnaires and tests. For these students, this assignment was a formal part of the 7-credit course Educational Technology at Ghent University. The 46 master students were divided over the 19 classes participating in this study. There were several reasons for designing the study in this way. First, the classroom teachers did not have the time to go through a training period and the interventions had to be carried out according to a set of instructional principles. Second, the master students had more expertise in the theoretical backgrounds of CSCL and were more prepared and familiar with the inquiry-based learning environment. Third, master students acted as a researcher as well. After a specific training, master students were responsible for data collection through a semi-structured interview, pre- and posttests, and observations during the sessions.

The inquiry-based learning environment used in this study was the Web-Based Inquiry Science Environment (WISE). WISE is developed to provide a solid technology platform that allows teachers to adopt new forms of inquiry-based instruction (Slotta & Linn, 2009). Based on Peters & Slotta (2009) a Flemish WISE-curriculum project was co-designed in partnership with science teachers and technology specialists. The project addressed global warming and climate change and was connected to the standards-based curricula in secondary education. The intervention was developed taking into account research-based design principles promoting knowledge integration and scientific inquiry.

During the first session, secondary students completed the individual pretest and were introduced to WISE. Depending on the time left, they also started in dyads the first activity of the project. The total project consisted of four main activities. Students worked in the same dyads during the whole intervention and navigated through the sequence of inquiry activities using the inquiry map in the WISE environment. During the project, students were asked to write their answers down in reflection notes. Consequently, all of their project work was stored in a database, which was accessible to teachers and researchers for purposes of assessment. Finally, all students completed the individual posttest.

During the sessions, master students were asked to act as a “leader from the within” instead of a “guide on the side”. A “leader from the within” does not only monitor students but actively engage the students, helps them to synthesize their views, and maintains a dynamic process of exchange within the classroom (Slotta & Linn, 2009). After each session master students provided electronic, both positive and critical, feedback through the feedback tool of WISE.

## Data Collection, Instruments and Analysis

In this study data was collected using a mixed method approach.

First, a pretest-posttest design was used to assess the impact of the intervention. The test consisted of three main parts which assessed content knowledge, the ability to do scientific inquiry, and the attitude and engagement towards science. In the first part, thirteen assessment items were developed to measure content knowledge, eight of them were knowledge/multiple-choice items and five of them were explanation items that went beyond reproducing knowledge and asked students to connect ideas in arguments. The latter items were scored using the knowledge integration rubric that rewards both accurate and connected ideas, created by TELS. This rubric contains a number of proficiency levels. The higher the proficiency level, the more complex the skills that the students have to master to tackle the scientific problems. In the second part, students were asked to read a research article and to generate the underlying hypotheses and research questions. In order to assess students’ attitude and engagement towards science, in the third part, three scales of the international PISA (2006) questionnaire were used: support for science, interest in science and responsibility towards resources and environments. Response options consisted of a 5-point Likert scale.

Second, data was obtained through content analysis of all students’ reflection notes during the four successive science lessons and the additional observation reports made by the master students.

Finally, after completing the project, semi-structured interviews of the classrooms teachers were conducted to assess their acceptance and attitude towards the innovative environment and approach. Furthermore, focus groups were organized twice with the master students to reflect on their experiences with the organization and implementation of the web-based sessions. Special emphasis was put on their role as a teacher in a web-based collaborative project.

Data analyses were conducted using the software program SPSS, version 15.0. Paired sample t-tests, independent sample t-tests and analysis of variances were conducted to determine the impact of the intervention.

# Results

## General results

Based on the observations and focus groups, we found that master students had difficulties to perform as a "leader from the within". Most of the time they monitored the way students were progressing and mainly responded to any confusion, in this way they acted generally as a "guide on the side". They were quite occupied managing the technology and monitoring students. Secondary school students who did not ask for help or did not have obvious technical problems that required help, were not stimulated to do more. Master students asked on a regular base how students were doing but when these students responded positive, no follow up questions were asked.

Regarding the classroom teachers, all of them were positive about this experience. They were especially pleased to have their students learn in a web-based collaborative inquiry environment and are willing to participate in future research.

## Content knowledge, inquiry skill and attitude and engagement towards science

Based on the pre- and posttest items we can determine that, overall, there is significant improvement in terms of knowledge and explanation. As shown in Table 1, students made progress in connecting ideas in arguments (explanation items) and their active knowledge (content knowledge items) about climate change has increased. In the pretest the numbers of incorrect and irrelevant answers were significant higher and knowledge about the topic was rather isolated.

Table 1: Average item mean, average item standard deviation, and differences (effect size and t-test for paired samples) between pretest and posttest using the individual as the unit of analysis

| | Pretest | | Posttest | | Difference | |
|---|---|---|---|---|---|---|
| Different scales | M | SD | M | SD | ES | T (df) |
| Content knowledge items | 1.38 | 0.57 | 2.23 | 0.44 | 1.6 | 27.97 (364)** |
| Explanation items | 1.70 | 0.68 | 2.62 | 0.66 | 1.36 | 23.93 (365)** |
| Inquiry skill | 1.8 | 1.03 | 2.17 | 0.82 | 0.41 | 6.49 (357)** |
| Support for science | 3.69 | 0.43 | 3.68 | 0.48 | | -0.237 (327) |
| Interest in science | 3.43 | 0.68 | 3.47 | 0.67 | 0.05 | 2.02 (314)* |
| Responsibility towards resources and environments | 4.07 | 0.48 | 4.13 | 0.49 | 0.12 | 2.95 (305)** |

* $p < 0.05$ ** $p < 0.01$
N=375

The results for the second part of the test where students were asked to deal with a research article are also significantly higher. Students became better in generating the underlying research questions and hypotheses in a research article. However, the qualitative data based on the reflection notes and observation reports force us to refine this result. We observed frequent use of the 'copy and paste' function, so that text fragments from the internet were directly included in their answers. Furthermore, students tend to reduce the whole task to finding an obvious answer on a particular website while less attention was paid to understanding the content and critical thinking.

With regard to the impact on attitude and engagement towards science, the results are less consistent. Concerning the support for science, we noticed that this is quite high at the pretest (3.6 on 5-point-likert), however no significant differences can be found compared to the posttest. However, the other two attitude scales (interest in science and responsibility towards resources and environments) do show significant differences between the pre- and the posttest although the effect sizes are low.

## Gender differences

Boys and girls do not score significantly different on the knowledge part of the pretest. In contrast, the scores on the posttest do significantly differ. Girls outperformed boys on knowledge items ($t$ = 3.09, $df$ = 367, $p$ < .01) as well as on explanation items ($t = 4.32$, $df = 368$, $p < .01$). In other words, girls benefited more than boys from the intervention. With respect to inquiry skills, no significant gender differences can be revealed.

With regard to the attitude towards science, no significant gender differences are found concerning the *support for science*, this in contrast to the *interest in science* and the *responsibility towards resources and environments* where significant gender differences are revealed. Girls had a significant lower interest in science on the pretest, but this difference disappeared after the intervention. Regarding the responsibility towards resources and environments, we can observe the same trend as in the knowledge part. Boys and girls do not score significantly different on the pretest, but in the posttest girls were more positive than boys on this attitude scale ($t = -2.36$, $df = 269$, $p = .02$).

## Discussion and conclusion

This study aimed to get students engaged with science content through inquiry and action. More specifically, this study dealt with the use and impact of web-based collaborative inquiry as a promising approach for secondary science education on students' understanding of scientific phenomena, their inquiry skills, and their attitude and engagement towards science. In addition, this study examined gender differences.

The present study demonstrated the effectiveness of web-based collaborative inquiry. This effectiveness is especially shown for girls. On the posttest, girls outperformed boys on knowledge items, explanation items, and responsibility towards resources and environments. This implies that this innovative instructional approach is an effective contribution in the attempt of making science accessible and interesting to all and to rectify the gender imbalance in science education (TIMMS, 2003; PISA, 2006). Overall, after the intervention, students made significant gains in understanding standards-based science concepts and improved on the inquiry test. Yet, we can conclude that information seeking as a part of an inquiry-based learning environment seems to be a complex and difficult progress for these students. During the web-based inquiry project students were faced with information problems, tasks that require them to identify information needs, locate corresponding information sources, extract and organize relevant information from each source, and synthesize information from a variety of sources (Brand-Gruwel, Wopereis, & Vermetten, 2005). It is often assumed that students naturally master this complex cognitive skill of information problem solving. According to this study, this assumption can be countered. According to previous research, we know that metacognitive knowledge and awareness is critical for students to be able to control and regulate their information problem-solving processes (Lazonder, 2003). However, little is known so far about effective instruction in information problem solving. Further research planned in March 2010 will provide insight into effective scaffolds to support student-directed inquiry. Furthermore, future research needs to focus on what teachers can do in an online classroom to effectively support student learning at more than a merely technical level, given that CSCL works best if the teacher interacts with the students and maintains a dynamic process of exchange within the classroom (Slotta & Linn, 2009). Unfortunately, our findings show that most of the time, the teachers acted mainly as a "guide on the side".

## References


Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. Paper presented at the AERA 2003 Symposium, Chicago, IL.

Bransford, J. D., Brown, A.L., & Cocking, R. (1999). How people learn: Brain, mind, experience, and school. Washington, D.C.: National Academy Press.

Collins, A. (1986). A sample dialogue based on a theory of inquiry teaching (Tech. Rep. No. 367). Cambridge, MA: Bolt, Beranek, and Newman

Cox, M., Abbott, C., Webb, M., Blakely, B., Beauchamp, T., & Rhodes, V. (2004). ICT and pedagogy: a review of the research literature. ICT in Schools Research and Evaluation Series, 18. London: DfES/BECTA.

DeBoer, G. E. (1991). A history of ideas in science education. New York: Teachers College Press.

De Meyer, I. (2008). Science competencies for the future in Flanders. The first results from PISA 2006. Ghent: OECD.

European Commission.(2008). Flash Eurobarometer: Young people and science: analytical report, from http://ec.europa.eu/public_opinion/flash/fl_239_en.pdf

Flemish Governmental enquiry (2005). Say what you want…about sciences, from http://ewi-vlaanderen.be/documenten/rapport_resultaten_enquete.pdf

Galili, I. (1996). Students' conceptual change in geometrical optics. International Journal of Science Education, 18, 847–869.

Gillies, R.M. (2003). Structuring cooperative group work in classrooms. International Journal of Educational Research, 39, 35-49.

Kyle, W. C., Jr., et al. (1985). What research says: Science through discovery: students love it. Science and Children, 23(2), 39-41.

Lazonder, A. W. (2003). Principles for designing web searching instruction. Education and Information Technologies, 8, 179-193.

Lindberg, D. H. (1990). What goes 'round comes 'round doing science. Childhood Education, 67(2), 79-81.

Lou, Y., Abrami, P.C., & d'Appolonia, S. (2001). Small group and individual learning with technology: a meta-analysis. Review of Educational Research, 71(3), 449-521.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004), Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Nichols, J.D. & Miller, R.B. (1994). Cooperative Learning and Student Motivation. Contemporary Educational Psychology, 19, 167-178.

Nichols, J.D. (1996). The Effects of Cooperative Learning on Student Achievement and Motivation in a High School Geometry Class. Contemporary Educational Psychology, 21, 467-476.

Osborne, M., & Freyberg, P. (1985). *Learning in science: Implications of children's knowledge*. Auckland, New Zealand: Heinemann.

Peters, V. L., & Slotta, J. D. (2009). *Co-designing curricula to promote collaborative knowledge construction in secondary school science*. Proceedings of the 9th international conference on Computer supported collaborative learning - Volume 1, Rhodes, Greece.

Rakow, S. J. (1986). Teaching science as inquiry. Fastback 246. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuiLE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), Cognition and instruction: Twenty-five years of progress (pp. 263-305). Mahwah, NJ: Erlbaum.

Sedeno, E. P., Balmaseda, I. S., Suarez, M. J. M., & Dauder, S. G. (2008). Perception of Science and Technology in teenage students from Madrid. *Arbor-Ciencia Pensamiento Y Cultura, 184*(733), 949-966.

Slotta, J. D. (2004). Web-Based Inquiry Science Environment. In M. C. Linn, E. A. Davis & P. Bell (Eds.). Internet Environments for Science Education (pp. 203-231). Mahwah, NJ: Lawrence Erlbaum Associates.

Slotta, J. D., & Linn, M. (2009). WISE Science: Web-Based Inquiry in the Classroom. New York: Teachers College Press.

Springer, L., Stanne, M. & Donovan, S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. Review of Educational Research, 69, 21-51.

Stark, R., & Gray, D. (1999). Gender preferences in learning science. International Journal of Science Education, 21(6), 633–643.

Stevens, R.J. & Slavin, R.E. (1995). The Cooperative Elementary School: Effects on Students' Achievement, Attitudes and Social Relations. American Educational Research Journal, 32, 321-352.

Tondeur, J., van Braak, J., & Valcke, M. (2007). Towards a typology of computer use in primary education. Journal for Computer Assisted Learning, 23(3), 197-206.

van Joolingen, W. R., de Jong, T., Lazonder, A. W., Savelsbergh, E., & Manlove, S. (2005). Co-Lab: Research and development of an on-line learning environment for collaborative scientific discovery learning. Computers in Human Behavior, 21, 671-688.

Wallace, R. M., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the web: Students online in a sixth-grade classroom. Journal of the Learning Sciences, 9(1), 75-104.

Woodgate, D. & Stanton Fraser, D. (2007). Workshop of Emerging Technologies for Inquiry-Based Learning in Science. Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education, from http://aied.inf.ed.ac.uk/AIED2007/InquiryBasedLearning.pdf

# Reconceptualizing Mathematical Learning Disabilities: A Diagnostic Case Study

Katherine E. Lewis, University of California, Berkeley, kelewis@berkeley.edu

**Abstract:** Mathematical learning disabilities (MLDs) research aims to define, diagnose, and remediate the cognitive impairments that characterize MLDs. Unfortunately little progress has been made towards these goals, because of unresolved methodological issues involved in distinguishing general low math achievement from low math achievement due to MLDs. To address these methodological issues, I conducted a small-scale exploratory study of MLD, which relied on several data sources to ensure that a student's low achievement was due to persistent and debilitating difficulties with mathematics that cannot be attributed to non-cognitive sources. I conducted four weekly video-taped one-on-one tutoring sessions with students focused on the topic of fractions. In this paper I present a case study of one student with MLD. Analysis indicated that the student understood mathematical representations of fractional quantity in atypical ways and this alternative understanding was problematic for the development of more advanced fraction concepts.

## Introduction

Lisa counts on her fingers to solve addition problems, cannot add fractions, and has just failed an arithmetic course. Lisa is a 19-year-old college student. Despite her best efforts and years of tutoring, she has not mastered elementary mathematics. Unfortunately, research on mathematical learning disabilities (MLDs) offers little insight to help understand why students, like Lisa, experience persistent, pervasive, and debilitating difficulties with mathematics. How are the difficulties experienced by students with MLD qualitatively different than those experienced by all students? What are the sources of the student's difficulties? How can researchers distinguish students with MLDs from students with low mathematics achievement? These difficult questions remain unanswered and are the focus of this exploratory study.

Research on MLD is in its infancy, as compared to research on more general learning disabilities (Fletcher, Lyon, Fuchs, & Barnes, 2008; Mazzocco 2007). Learning disability research has traditionally focused on language-based learning problems, which involve difficulties with processing phonemes (the smallest unit of sound), and therefore provides little insight into the cognitive difficulties students experience in mathematics (Fletcher et al., 2008). MLDs are typically understood as core cognitive deficits, which result in a student's failure to learn or remember mathematical procedures or concepts (Geary, 2004). However, there is currently no consensus operational definition of MLDs and no diagnostic instruments available to accurately identify students with MLDs (Murphy, Mazzocco, Hanich, & Early, 2007). Consequently, researchers rely on low math achievement test scores to classify students, which does not address the myriad reasons, besides MLDs, a student could have performed poorly on the test. Given the definitional and methodological issues hindering this field, I adopt an alternative theoretical and methodological approach to the study of MLDs.

In my research I adopt a *difference*, rather than *deficit*, understanding of disability and leverage advances made in language-based learning disability research for both subject identification and analytic foci. I ensure that the students, not only meet the low-achievement criteria established by other researchers, but that the student's low-achievement is likely due to a cognitive rather than environmental factor. I define MLD by building upon innovative subject classification methodologies used in learning disability research, which require that students demonstrate (1) low achievement (for my purposes, in mathematics), (2) no evidence of confounding factors explaining that low achievement, and (3) failure to respond to an instructional intervention (Fletcher et al., 2008). In this case study, I conducted a detailed qualitative analysis of one student as she engaged in the process of learning over several tutoring sessions with a particular focus on the student's ability to conceptualize, represent, and work with numbers, specifically fractions. Just as the capacity to process phonemes is fundamental to reading competency, numbers are the fundamental building block of mathematics (Dehaene 1997, Landerl, Bevan & Butterworth, 2004). Numbers, or quantities, get represented in a variety of modalities (e.g., orally: "three-fourths", numerically "3/4", and graphically: area model for 3/4, see Figure 1). Fluency with these representational systems is a crucial part of what it means to do and understand mathematics (Ball, 1993; Lesh, Post, & Behr, 1987; NCTM Principles and Standards, 2000). This suggests that students' fluency with number, quantity, and representations is a crucial component of mathematical development and a productive analytic lens through which to explore MLDs.

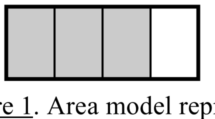


Figure 1. Area model representation of 3/4.

This exploratory case study focuses on identifying characteristics of MLDs. I conducted detailed diagnostic analyses of tutoring sessions focusing on the difficulties the student experienced, the kinds of atypical understandings she demonstrated, and her use of representations. This research addresses the following questions:

1) What is the nature of the student's difficulties during the tutoring sessions?
2) What are the characteristics of the student's atypical understanding of mathematical concepts over the course of the tutoring sessions?
3) How does the student use and understand mathematical representations?

Analysis indicated that the student experiences persistent difficulties with fraction equivalence and operations. Additionally, the student demonstrated atypical understandings of representations of fractional quantity, which hindered her ability to make sense of more complex fraction concepts.

## Prior Research on Mathematical Learning Disabilities

Prior research on MLDs has made limited progress towards understanding this complex phenomenon due to a lack of a consensus definition for what constitutes a MLD. Currently, it is a widely accepted practice to use achievement test scores below a given cut-off (typically the 25$^{th}$ percentile) to identify students with MLDs (Murphy et al., 2007; Geary, 2004). Various critiques of this approach have been identified. The cut-off score used is arbitrarily selected (Francis et al., 2005), and can range from the 8$^{th}$ percentile to the 46$^{th}$ percentile for studies of MLDs (Gersten, Clarke, & Mazzocco, 2007; Swanson & Jerman, 2006). This has led to variability in the profile of students classified as having MLDs, suggesting researchers are not all studying a common phenomenon (Murphy et al., 2007). In addition, researchers rarely attempt to control for factors that are well-established as correlated with low-achievement. Hanich, Jordan, Kaplan, and Dick (2001) report that studies of MLDs have an over-representation of minority, poor, and non-native English speaking students in the MLD group. This signifies that the proxy of low achievement does not adequately address the complex social components at play in the operational definition of MLDs. Given the lack of adequate subject identification methods, it can be argued that researchers are studying *general characteristics of low math achievement rather than MLDs*. To address this methodological hurdle, I propose adopting an alternative theoretical perspective and methodological approach.

## Theoretical Approach

Rather than adopting the traditional conceptualization of MLDs as cognitive *deficits* (Geary, 2004), I conceptualize MLDs in terms of cognitive *difference.* A deficit understanding of MLDs orients researchers to identifying the ways in which a group of students classified as having MLDs are in some way deficient when compared to a typically achieving group. Therefore, any statistically significant group difference is assumed to be consequential and indicative of the students' MLDs. In contrast, a *difference* understanding of MLDs orients researchers to identifying the nature of the individual student's qualitative differences. Instead of asking: "by how much is the MLD group deficient?" this approach asks: "what differences are consequential for an individual with a MLD and how are they consequential?" Recent MLD research supports the idea that qualitative differences exhibited by students with MLDs may ultimately be the key to differentiating these learning disabilities from low math achievement (Mazzocco, Devlin, & McKenney, 2008). Therefore, I adopt a difference model of MLDs, building on a Vygotskian perspective of disability, in which a student with a disability "is not simply a child less developed than his peers but is a child who has developed differently" (Vygotsky, Knox, Stevens, Rieber, & Carton, 1993 p. 30). I argue that the nature of a student's disability is revealed through an analysis of *qualitative* differences in how the student is reasoning over time, rather than in a quantification of deficits in knowledge measured at one point in time.

## Fractions

For a number of reasons, fractions provide a mathematically rich terrain to explore how students with MLDs make sense of mathematical concepts. First, MLDs research to date has focused almost exclusively on students' difficulties remembering math facts (e.g., 4+5=9; Gersten, Jordan, & Flojo, 2005; Swanson & Jerman, 2006). Equating MLDs with difficulty learning math facts reduces mathematical cognition to efficient and accurate production of an answer. This finding does little to explain the range of difficulties students may have across the variety of mathematical domains (Dowker, 2005; Geary, 2005). Second, the kinds of difficulties that typically achieving students experience when learning fraction concepts have been extensively documented in math cognition research (see Lamon, 2007 for a review). It is therefore possible to classify the difficulties demonstrated by students with MLDs as typical or atypical. Considering MLDs in the context of fractions allows for the exploration of procedural, conceptual, and representational difficulties that students with a MLD may experience in this more complex mathematical domain.

## Methods

This case study was drawn from a larger study that involved an extensive subject recruitment and classification process to ensure that each student's mathematical difficulties were due to an MLD and not other factors. Low achievement in math was considered a necessary but not a sufficient criterion for a MLD classification. In addition, subjects classified as having MLDs did not exhibit any evidence of other factors that could explain their low achievement and when given a sufficient instructional intervention, the student failed to show learning gains. Figure 2 provides an overview of each phase of this project including: initial recruitment, data collection, subject classification, and diagnostic analysis. Each will be discussed in turn.

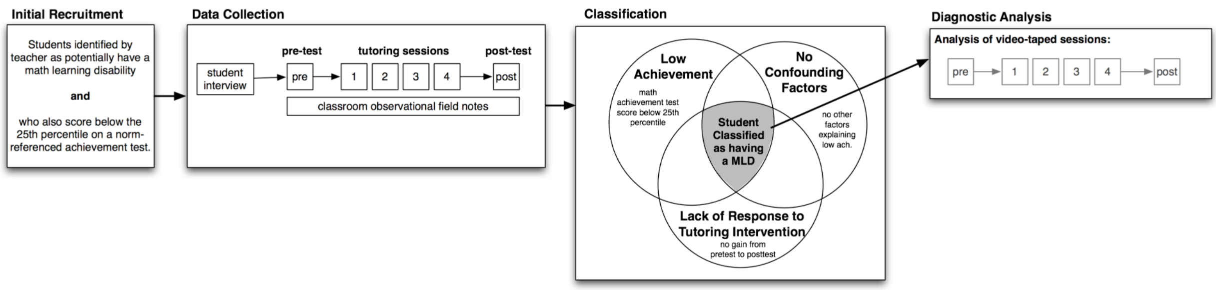



Figure 2. Overview of the design of the study.

### Initial Recruitment

Subjects were recruited from several school sites, including: a public middle school, a private high school, and a community college. Math teachers from these schools were asked to identify students who they believe may have a MLD. For those students identified, only students who met the canonical MLDs low-achievement qualification (scoring below the $25^{th}$ percentile on their most recent math achievement test) were invited to participate in the study.

### Data Collection

#### Student Interview

A 30-minute interview was conducted with the student before the administration of the pretest. The interview questions attempted to elicit the student's characterization of his/her difficulties (e.g., what about learning math is hard for you?), perceived level of effort (e.g., do you tend to complete all your math homework?) and available resources (e.g., what do you do when you don't know how to answer a math problem?)

#### Pretest / Posttest

A videotaped semi-structured clinical interview-style pretest and identical posttest was administered to all subjects. The assessment was designed to cover all fraction concepts targeted in the tutoring sequence (see Table 1 for exemplar problems). The pretest was administered the week before the commencement of tutoring sessions and the posttest was administered the week following the completion of the tutoring sessions. The student's answers and explanations were scored according to a rubric and total scores were calculated for each of the target concepts.

#### Tutoring Sessions

Four hour-long weekly videotaped tutoring sessions were conducted with the subjects focusing on fraction concepts. These sessions were designed based on prior research on the teaching and learning of fraction concepts and have been piloted and refined. In each session a sequence of problems was posed to the student, each problem was intended to build upon prior problems and provoke conversations about important math content. Table 1 provides details of the targeted concepts and exemplar problems from the pretest/posttest and tutoring session problems. Each of the problems had follow-up questions and/or counter suggestions. To refine the tutoring protocol (pretest, tutoring sequence, and posttest), it was administered to fifth-grade students with no learning difficulties. The fifth-grade students demonstrated substantial gains from pretest to posttest and therefore, this instructional tutoring sequence was considered to be an adequate learning environment.

Table 1: Target fraction concepts, exemplar problems from pretest/posttest and tutoring sessions, and references for related research.

| Concept | Details | Exemplar Pretest/Posttest Problem | Exemplar Tutoring Session Problem | Related Research |
|---|---|---|---|---|
| Fraction Representation and Interpretation | A fraction represents a single number rather than two independent numbers, which involves determining the number of equal parts of a unit out of the total number of equal parts. | Can you circle the pictures that you think are the same as 1/2? [*representations include number lines, discrete pieces, area models. Both 1/2 and equivalent fractions (e.g. 2/4) are included along with common student errors.]* | Without writing down any numbers can, you draw a picture of 3/4 so that other people would know that it's a picture of 3/4? | Ball (1993)<br>Lesh et al. (1987)<br>Mack (1990)<br>Post et al. (1985)<br>Behr et al. (1992)<br>Saxe et al. (2005) |
| Fraction Comparison | Fraction size depends on relationship between two numbers. | Which is bigger or are they equal?<br>[*2 area models are presented based on Armstrong & Larson*] | Put these cards in order from least to greatest. (Card values: 0, 1/1000, 1/2, 2/3, 8/12, 9/10, 7/7, 1, 1 2/5, 3/2) | Armstrong & Larson (1995)<br>Post et al., (1985) |
| Equivalent Fractions | Many different fraction names designate the same amount | Can you come up with a fraction equivalent to: ½, 1/3, 2/5, 8/12? | Using a whole sheet of paper, draw 2/3 of a cake. Figure out how you can cut the cake again so you still have even pieces. What has change and what has stayed the same? | Lamon (1996)<br>Lamon (2006)<br>Lamon (2007)<br>Post et al. (1985)<br>Ni (2001)<br>Kamii & Clark(1995) |
| Fraction Addition | Addition and subtraction of fractions requires common denominators | 1/2+1/4= | 1/2+1/3 =<br>How would you solve this problem using pictures? | Mack (1992)<br>Steffe (2003) |

## Classroom Observations

Classroom observations were conducted concurrently with the tutoring sessions. Field notes documented any evidence of the student's attention or behavior issues, level of effort, and performance relative to peers. This was an exclusionary data source intended to capture evidence of students who have non-cognitive issues, which could explain their poor performance.

## Subject Classification

Several data sources were drawn upon to determine if the student's low achievement could be explained by other factors. Those factors highly correlated with low math achievement were considered confounding factors for the purposes of this study. These included: attention or behavior problems, lack of English fluency, insufficient resources, math anxiety (Ashcraft, Krause, & Hopko, 2007; Chatterji, 2005; Diversity in Mathematics Education, 2007; Zentall, 2007). Classroom observations, interviews, and the videotaped pretest, tutoring sessions, and posttest were used to determine if any students exhibited or reported potential confounding factors. I do not assume that MLD cannot co-occur with these factors, but that disentangling the effect of the confounding factor from the effect of MLD is beyond the scope of this study, and therefore any students demonstrating these factors were excluded. The goal was not to identify all students with MLDs, but to ensure that all students included in the larger study had an MLD. In addition, this tutoring protocol attempts to rule out the possibility that the student's history of low-achievement was simply due to poor instruction. If the student *did* show gains from pretest to posttest, it suggests that prior instruction may have been insufficient and an underlying cause of the student's low achievement. Conversely, if the student *did not* show gains from pretest to posttest, this suggests that poor instruction was not the primary cause of the student's low achievement and indicates that the student likely has an MLD. For example, Figure 3 displays a comparison of a fifth grade student's pretest-posttest change as compared to that of a community college case study student (Lisa), who I classify as having a MLD. The contrast between the student's scores (as measured by slope from pretest to posttest) is striking, providing strong substantiation of an MLD classification for Lisa.

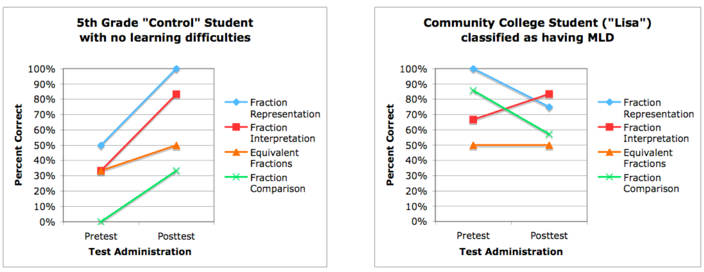



Figure 3. Comparison of a pretest/posttest scores for a typically-achieving fifth grade student and a community college student who is classified a having a MLD.

## Diagnostic Analytic Approach

All videotapes of the sessions (pretest, tutoring sequence, and posttest) were transcribed and all student work was scanned. A microgenetic bottom-up analysis of the whole data corpus for each student was conducted (see Schoenfeld, Smith, & Arcavi, 1993, for an example of this kind of analysis applied to graphing). This kind of analysis involved iterative passes through the data in an attempt to generate analytic categories that capture the nature of the student's understanding (including difficulties, atypical understandings, and uses of representations). The generation and refinement of the analytic categories involved a process of compiling the instances of a proposed kind of understanding, specifying inclusion and exclusion criteria, generating a hypothesis about the understanding, and returning to the entire data corpus (both transcripts and videos) to support or reject the hypothesis. Analysis focused on identifying the kinds of student understanding that resulted in difficulties or appeared to be consequential for how the student was making sense of the mathematics. Some of these student understandings were well documented as common problems in the math cognition literature (e.g., thinking that 1/100 is larger than 1/10 because 100 is larger than 10) and were therefore not the focus of my analysis. The rationale for focusing on these atypical understandings was an attempt to evaluate what, beyond the canonical difficulties, resulted in the student's failure to learn during the tutoring sessions.

## Results

The focal case study subject, Lisa, was considered to have a MLD because she exhibits low achievement, which was not attributable to other factors, and because she failed to respond to the tutoring instruction. Lisa had just completed her first year of community college, and her math achievement test score was in the lowest 25th percentile. Interviews and observations revealed that there were no identifiable confounding factors; Lisa was a white upper-middle class, native English speaker with no attention or behavior problems. Lisa's overall score did not improve from prettest to posttest, and for some fraction topics, her score was lower at the time of the posttest (see Figure 3).

### Diagnostic Analysis

Why does Lisa fail to learn from the tutoring sessions? A detailed analysis of all sessions (pretest, tutoring sessions, and posttest) indicates that Lisa experienced difficulties with comparing fractions, adding fractions, and generating equivalent fractions throughout all of the sessions including the posttest. Her understanding of fractional quantity appeared to be unstable, in that she operated on representations of fractions as if they represented a partitioning or removal activity as opposed to a fractional quantity. This alternative understanding of fraction representations appears to derail Lisa's ability to build an understanding of fraction equivalence, comparisons, and operations.

A detailed analysis of Lisa's use and understanding of area model representations was conducted. Area models are a common pedagogical representation for fractions, in which a shape is partitioned into evenly-sized pieces corresponding to the denominator, and then the number of pieces corresponding to the numerator is shaded. For example, an area model for 7/12 would be drawn by partitioning a shape into 12 even pieces and then shading 7 of those pieces. Lisa, however, understood area models as if they signified an *action* as opposed to a *quantity*. This understanding has two primary manifestations. First, the shading of area models was understood as the amount taken away, rather than indicative of the fractional quantity. Second, Lisa understood "one-half" as an *action* of halving rather than the quantity of 1/2. These understandings were considered atypical because they have not been documented in the math cognition literature, and were not experienced by any of the fifth-grade subjects used to validate the tutoring protocol. In addition, this alternative understanding proved to be highly consequential for the student's ability to make sense of more complex fraction concepts.

#### Understanding Area Models as Representing an Amount "Gone"

Lisa understood the shading of area models as the amount removed, rather than indicative of the fractional quantity. Lisa's focus on the removal action as opposed to quantity led to inconsistently between her construction and interpretation of area model representations. In the following example, she was asked to compare the fractions 7/12 and 1/2. She correctly constructed a drawing of 7/12 by partitioning the shape into 12 pieces and shading 7 of those pieces. However, she incorrectly determined that 7/12 is *smaller* than 1/2 and when I asked her to explain she referred to the shaded pieces as "gone" and began attending to the five pieces "left."

a)  b) 

Figure 4. Reproduction of the student's drawing of 7/12, in which she (a) partitioned a shape into 12 pieces, and then (b) colored 7 of the pieces.

L: I mean, ok, so let's say that this is the cake (gestures back and forth over entire shape– Figure 4b) and seven pieces are gone (makes sweeping motion over the shaded pieces of the shape).

…

K: Which one do I have more for?

L: Um. The half.

Although she drew a canonical area model (using shading to represent the numerator) she subsequently interpreted the shaded region of area models as being "gone". This disconnection between her correct construction and incorrect interpretation was not easily remediated and persisted throughout the tutoring sessions to the posttest. I argue that, for Lisa, the shading represented the *action* of removal rather than the fractional *quantity*. In this example, her ability to make sense of the fraction comparison task was compromised by her atypical conception of the mathematical representation. Throughout the tutoring sessions, this understanding of fractional representations in terms of action appeared to be highly consequential and detrimental to her ability to build a more complete understanding of fraction equivalence, fraction comparison, and fraction addition.

## Understanding "one-half" as Partitioning Rather Than a Quantity

Lisa often represented the fraction one-half as a partition rather than a quantity. For example, in the posttest she drew one-half by "halving" the shapes but not by designating a quantity of one-half (see Figure 5). When asked about what part of the drawing was one-half, she pointed to or redrew the line darker.

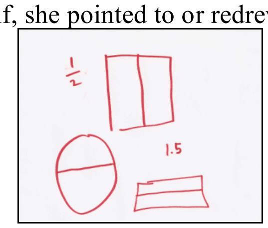



Figure 5. The written work of the student's answer to the posttest item "draw or write one-half".

This action orientation to area model representations of quantity appeared to be consequential to her inability to develop a robust understanding of equivalent fraction and fraction addition concepts. For example, at the end of one tutoring session she was attempting to write in her journal about what she had learned. During the tutoring session she had been correctly generating equivalent fractions by repartitioning area models (see Figure 6).

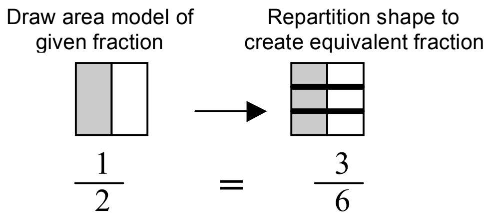



Figure 6. Illustration of the process used to create equivalent fractions during the tutoring session.

Although she had been successfully using area models to create equivalent fractions during this session, when she attempted to summarize what she had learned with an example, she incorrectly determined that 1/2 equaled 1/6. She began by drawing a square and partitioning it into two pieces (without shading). She then added two new horizontal partitions. However, because her drawing did not involve shading of 1/2, she incorrectly determined that 1/2 was equivalent to 1/6.

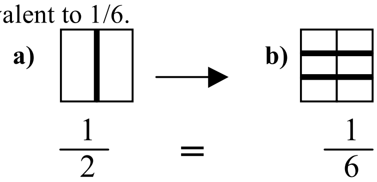



Figure 7. Recreation of Lisa's attempt to construct an equivalent fraction for 1/2 which omitted shading.

L: Ok, so like, what we were doing, if you have like one-half (draws shape and draws vertical line, see Figure 7a), and then like, cut it like that, (draws partitions, see figure 7b) it stll stays the same. It is just cut into different sections then.

K: So, how many sections is it cut into?

L: Six. So one-sixth?

After representing 1/2 as the partitioning line rather than by shading one-half of the shape, she produced a repartitioned area model with no shading. She correctly determined that the shape was divided into six pieces and guessed that the equivalent fraction for 1/2 was 1/6. This recording of an example, intended to be the culminating activity for this tutoring session, was undermined by her unconventional representation of one-half as a partition rather than a quantity. Lisa's answers and gestures throughout the tutoring sessions suggested that she understood one-half to be represented by the act of partitioning rather than a resulting quantity. Lisa's understanding of both the shaded region as signifying an amount taken away and her understanding of one-half as a partition, involved what proved to be a detrimental orientation to the area model representation as *action* rather than a *quantity*.

## Conclusion

Lisa's atypical understanding, representation, and manipulation of quantities appeared to be a fundamental source of her difficulties, which interfered with her ability to develop a foundational fraction concepts of equivalence and fraction operations. The two primary manifestations: understanding shading as taken away and understanding one-half as a partition as opposed to a quantity, have appeared in two other students in the larger study who are classified as having MLD. This suggests that the atypical understanding of representations of quantities may be a powerful finding and a productive avenue to consider in the design of diagnostic tools. Given this kind of diagnostic understanding of a student's difficulties, it is possible to begin designing and testing remediation tools, which help the student compensate for these difficulties. This case study explored the nature of MLDs, by identifying the kinds of problems a student with an MLD exhibits beyond the canonical problems typically experienced by students when learning mathematics. Future work should attempt to capitalize on the existence of these kinds of qualitative differences exhibited by students with MLD, rather than relying upon math achievement tests for classificatory purposes. Ultimately the goal of MLD research should not be to label students and classify their deficiencies, but to understand their differences and help them learn to compensate.

## References


Armstrong, B., & Larson, C. N. (1995). Students' use of part-whole and direct comparison strategies for comparing partitioned rectangles. *Journal for Research in Mathematics Education, 26*(1), 2-19.

Ashcraft, M. H., Krause, J. A., & Hopko, D. R. (2007). Is math anxiety a mathematical learning disability? In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is Math So Hard for Some Children?: The Nature and Origins of Mathematical Learning Difficulties and Disabilities* (pp. 329-348). Baltimore: Paul H Brookes.

Ball, D. L. (1993). Halves, pieces, and twoths: Constructing and using representational contexts in teaching fractions. In T. P. Carpenter, E. Fennema, & T.A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 157-195). Hillsdale, NJ: Erlbaum.

Behr, M. J., Harel, G., Post, T., Lesh, R. (1992). Rational number, ratio, and proportion. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296-333). New York: Macmillan Publishing.

Chatterji, M. (2005). Achievement gaps and correlates of early mathematics achievement: Evidence from the ECLS K–first grade sample. Education Policy Analysis Archives, 13(46).

Diversity in Mathematics Education (DiME) Center for Learning and Teaching (2007). Culture, race, power, and mathematics education. In F. K. Lester (Ed.) Second Handbook of Research on Mathematics Teaching and Learning (pp. 405-434). Charlotte, NC: Information Age Publishing, Inc.

Dowker, A. (2005). Early Identification and Intervention for Students With Mathematics Difficulties. *Journal of learning disabilities*, *38*(4), 324-332.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., Barnes, M. A. (2007). *Learning disabilities: From Identification to intervention.* New York, NY, US: Guilford Press.

Geary, D.C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4-15.

Geary, D. C. (2005). Role of Cognitive Theory in the Study of Learning Disability in Mathematics. *Journal of learning disabilities, 38*(4), 305-307.

Gersten, R., Clarke, B., & Mazzocco, M. M. M. (2007). Historical and Contemporary Perspectives on Mathematical Learning Disabilities. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is Math So Hard for Some Children?: The Nature and Origins of Mathematical Learning Difficulties and Disabilities* (pp. 7-27). Baltimore: Paul H Brookes.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early Identification and Interventions for Students With Mathematics Difficulties. *Journal of learning disabilities, 38*(4), 293-304.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance Across Different Areas of Mathematical Cognition in Children With Learning Difficulties. *Journal of educational psychology, 93*(3), 615-626.

Kamii, C., & Clark, F. B. (1995). Equivalent fractions: Their difficulty and educational implications. *The Journal of Mathematical Behavior, 14*(4), 365-378.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition, 93(*2). 99-125.

Lamon, S. J. (1996). The development of unitizing: Its role in children's partitioning strategies. *Journal for Research in Mathematics Education, 27*(2), 170-193.

Lamon, S. J. (2006). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Lamon, S. J. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In F.K. Lester, Jr. (Ed.) *Second handbook of research on mathematics teaching and learning* (pp. 629-668). Charlotte, NC: Information Age Pub Inc.

Lesh, R., Post, T., Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 33-40). Hillsdale, NJ: Erlbaum.

Mack, N. (1990). Learning fractions with understanding: Building on informal knowledge. *Journal for Research in Mathematics Education, 21,* 16-32.

Mazzocco, M. M. M. (2007). Defining and differentiating mathematical learning disabilities and difficulties. In D. B. Berch, & M. M. M. Mazzocco (Eds.) *Why is math so hard for some children?: The nature and origins of mathematical learning difficulties and disabilities* (pp. 29-48). Baltimore, MD: Paul H Brookes Publishing.

Mazzocco, M.M.M., Devlin, K.T. & McKenney, S. J. (2008). Is it a Fact? Timed Arithmetic Performance of Children With Mathematical Learning Disabilities (MLD) Varies as a Function of How MLD is Defined', *Developmental Neuropsychology, 33:*3, 318-344.

Murphy MM, Mazzocco MMM, Hanich L, & Early MC. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cut-off criterion used to define MLD. *Journal of Learning Disabilities.* 40:458-478

National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.

Ni, Y. (2001). Semantic domains of rational numbers and the acquisition of fraction equivalence. *Contemporary Educational Psychology, 26*(3), 400-417.

Post, T. R., Wachsmuth, I., Lesh, R., & Behr, M. J. (1985). Order and equivalence of rational numbers: A cognitive analysis. *Journal for Research in Mathematics Education, 16*(1), 18-36.

Saxe, G. B., Taylor, E. V., McIntosh, C., & Gearhart, M. (2005). Representing fractions with standard notation: A developmental analysis. *Journal for Research in Mathematics Education, 36*(2), 137-157.

Schoenfeld, A. H., Smith, J. P., Arcavi, A. (1993). Learning: The microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Ed) *Advances in Instructional Psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Steffe, L. P. (2003). Fractional commensurate, composition, and adding schemes learning trajectories of Jason and Laura: Grade 5. *The Journal of Mathematical Behavior. Special Issue: Fractions, Ratio and Proportional Reasoning, Part B, 22*(3), 237-295.

Swanson, H, L. (2007). Cognitive Aspects of Math Disabilities. In D. B. Berch, & M. M. M. Mazzocco (Eds.) *Why is math so hard for some children?: The nature and origins of mathematical learning difficulties and disabilities* (pp. 133-146). Baltimore, MD: Paul H Brookes Publishing.

Swanson, H. L., & Jerman, O. (2006). Math Disabilities: A Selective Meta-analysis of the Literature. *Review of Educational Research, 76*(2), 249-274.

Vygotsky, L. S., Knox, J. E., Stevens, C. B., Rieber, R. W., & Carton, A. S. (1993). *The collected works of L. S. Vygotsky, Vol. 2: The fundamentals of defectology (abnormal psychology and learning disabilities).* New York, NY, US: Plenum Press.

Zentall, S. S. (2007). Math performance of students with ADHD: Cognitive and behavioral contributors and interventions. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is Math So Hard for Some Children?: The Nature and Origins of Mathematical Learning Difficulties and Disabilities* (pp. 219-244). Baltimore: Paul H Brookes.

## Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education pre-doctoral training grant R305B090026 to the University of California, Berkeley. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.

# Sharing Educational Scenario Designs in Practitioner Communities

Astrid Wichmann, Jan Engler, H. Ulrich Hoppe
Collide Group, Department of Computer Science and Applied Cognitive Science,
University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany
{wichmann, engler, hoppe}@collide.info

**Abstract:** The possibility of sharing results and ideas is an important benefit of networked communities. In educational design, practitioners (e.g. teachers) specify educational scenarios that can be (re-)used, exchanged and modified at a later stage. In addition to supporting educational scenario design as such, our graphical editor SCY-SE offers functions to retrieve scenarios created by others based on a similarity measure. To examine the validity of this similarity measure in terms of correspondence with pre-defined cases, 25 participants were asked to produce graphical scenarios from given textual descriptions. First results strongly indicate that the calculated similarity was much higher between corresponding scenarios (related to the same text, but by different modellers) than in non-corresponding cases. Also, scenarios created by the same person show high dissimilarity for different cases. These results suggest that similarity-based matching is an effective and acceptable method to support the exchange among educational designers.

## Introduction

Networked communities of practitioners have a great potential for accumulating ideas and artefacts that can be (re-)used, exchanged and modified at a later stage. This view is consistent with Wenger's notion of communities of practice (Wenger, 1998). The ability to share and exchange ideas and/or artefacts needs a certain degree of awareness on the part of the participants, which includes awareness of other active members in the community and their interests (social awareness) as well as awareness of the objects or artefacts produced in the community (content awareness). In small local communities, social awareness and social interaction often guide the identification of objects of common interest. Yet, in realistically sized networked communities, this will only work to a very limited extent. Alternatively, objects of common interest have to be identified on the content level. Technically, this identification can be achieved through providing a similarity measure for pairs of objects. Here, we assume that a community repository will accumulate objects produced by community members, and that these objects can be seen as indicators of the current interests of their authors (at creation time). So, for each new object that is uploaded to the community repository we can automatically search for similar objects that already exist. These existing objects may be brought to the awareness of the new author, which may result in revising the content in regard of existing material or even in contacting the creators of the existing objects to coordinate further elaboration and other actions.

A similar approach within a community of science learners has been suggested by Hoppe et al. (2005). Here, the objects of interest were so-called "emerging learning objects" created by members of different local teams in an international learning community that shared different scientific challenges without personally knowing each other between the different local teams. Objects of common interest between two partners were also identified by similarity, exploiting the internal structure of the objects. The similarity measure had to work in the absence of user-defined metadata since it was assumed that a sufficiently complete and consistent indexing of emerging objects by the creators (high schools students) could not be expected. A comparable mechanism to identify congruence of interests in a community of scientists has been suggested by Francq & Delchambre (2005) based on a publication database with a similarity definition using automatic indexing techniques. Generally, similarity measures can be used as a source of continuous information on themes of personal interest. Similar functions are served by *recommender systems* (cf. Konstan & Riedl, 2002), which usually rely on information about usage processes, namely by computing object-object associations through co-occurrences in user traces.

In our study, we target a community of educational designers who produce semi-formal designs of structured learning scenarios using a graphical editing tool. The potential benefit of automatically identifying objects of common interests (i.e. specific educational designs) as a content awareness function is evident: It will be possible to relate new suggestions to existing ones early at design time. Thus, unnecessary redundancies can be avoided and synergy between similar approaches can be exploited. The question, however, is if the underlying similarity measure does really capture those features that would be considered as relevant by human judgment. Suggestions based on irrelevant or artificial similarities would soon lead to rejection of the support mechanism. To validate the similarity mechanism in this sense, we provided users with a number of different textually defined learning scenarios that were to be represented (or modelled) graphically using our editing tool. To be “valid” and acceptable the mechanism should adequately reproduce the original differences irrespective of personal preferences and idiosyncrasies. This makes another assumption, which is that the modelling task with the given tool itself is well-defined in that a substantial amount of relevant information can be transferred from the textual description to the graphical model. A negative outcome would still not clarify if the problem was with the similarity calculation or with the coding task (or both). A positive outcome, however, would confirm both the task-tool fit and the adequacy of the similarity measure. In the sequel, we will first describe the scientific context of this study in a European research project, then elaborate on the technical ingredients (editor and similarity measure) and finally report our empirical findings.

## Educational design of inquiry-based learning scenarios

The requirements for the work presented in this paper stem from a new European research project called SCY[1]. In the SCY project (de Jong, Joolingen van, Weinberger, 2009), students engage in inquiry learning activities supported by computer tools such as simulations and modelling software. In SCY-Lab (the SCY learning environment), students work on missions and meet challenges collaboratively and individually supported by learning material, tools and scaffolds. To meet the challenges of inquiry learning, the configuration of SCY-Lab is adaptive to learners’ capabilities and progress. In order to support the cyclic nature on inquiry learning, SCY-Lab provides tools and scaffolds, which provide just-in time support (Anjewierden, Chen, Wichmann, & Borkulo van, 2009). Another characteristic of inquiry learning is learners’ rich interaction with learning material. In SCY missions, students not only receive learning material (e.g. in form of text), they also produce material (e.g. in form of data by running experiments). The construction of artefacts that emerge from the learning process has been articulated as one of the central ideas in SCY. These emerging learning objects (ELOs) are re-usable and sharable products of learning activities, which are created by learners. As part of SCY’s pedagogical approach, several so-called *learning activity spaces* (LASs) have been identified. These are characterised by specific combinations of tools, activities and types of input/output objects. *Learning scenarios* are defined on an upper level by a set of LASs with partial sequencing.

One goal of SCY is to support a community of practitioners (teachers and instructional designers) in creating and sharing educational scenarios. Existing languages and tools in the area of learning design such as LAMS (Dalziel, 2003) or IMS/LD[2] appeared to be not sufficient for our purposes for different reasons: LAMS is too coarse grained with respect to capturing object-tool-activity relationships and results in quite fixed sequential models with too little flexibility. IMS/LD is more expressive with regard to tools, activities and roles, yet it lacks a conceptual model and graphical representation that would be usable by non-technical experts (cf. Miao, Harrer, Hoeksema, & Hoppe, 2007). Both approaches lack possibilities to adequately represent input and output relations with ELOs.

As argued above, effective community support has to include the sharing and re-use of previously developed scenarios. For teachers who develop new scenarios for SCY missions it is important to be able to find scenarios with specific characteristics. For instance, several strategies exist to introduce scientific texts to learners. One teacher might favour a concept mapping activity to identify and integrate new concepts while another teacher would request the students to summarize the text using a note-taking tool. Different approaches to facilitate sharing and re-use have been suggested. One approach is to select and filter scenarios based on certain categories such as age and subject. However, single categories like these may not be sufficient to capture what is relevant for the teacher (cf. Ronen, Kohen-Vacs, & Raz-Fogel, 2006). Another method is to let users (here: teachers) freely tag their design objects. Tagging, however, bears several problems. First, it is an additional effort, which is often avoided. Second, the degree of freedom and subjectivity in free tagging may lead to inconsistencies of various types, which will lead to unreliable results when used for retrieval. Therefore, we propose a similarity measure that reflects the two-level structure of our learning design and uses several content categories (such as activities, tools and object types). For instance, a teacher may want to focus on argumentative activities and might thus be interested specifically in other scenarios involving argumentation tools and ELO types compatible with argumentation.

## The SCY- SE Scenario Editor and graphical language

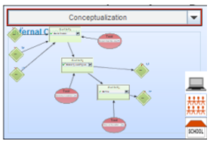
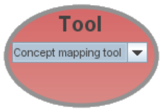
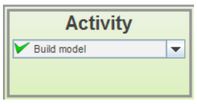
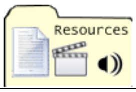
The SCY Scenario Editor (SCY-SE) has been designed and implemented following the definition of pedagogical requirements and the ensuing specification of a graphical representation language (de Jong et al., 2009). SCY-SE has been implemented as a plug-in to the multi-functional graphical modelling environment FreeStyler (Hoppe & Gaßner, 2002). FreeStyler provides a uniform platform for various visual languages with free-hand annotation

[1] SCY – “Science created by You” is an EU project of the 7th Framework Programme. For more information, see http://www.scy-net.eu (last visited in October 2009).

[2] For the full specification and more information in IMS-LD, see http://www.imsglobal.org/learningdesign (last visited in October 2009).

facilities and shared workspace facilities. The graphical user interface of SCY-SE offers a two-level view reflecting the pedagogical design: the scenario view combines several LASs with input and output ELOs, whereas the lower-level LAS view specifies activities, tools, intermediate ELOs and other resources. SCY-SE provides a toolbox ("palette") from which objects can be dragged onto the workspace.

Table 1. Concepts in SCY-SE

| Concept | | Specified by the SCY-SE user |
|---|---|---|
| ELO (Emerging Learning Object) |  | ELO characteristics (functional role, technical format, logical representation, belonging activity) |
| LAS (Learning Activity Space) |  | - LAS type (e.g., conceptualization, build, debate, experiment)<br>- Location (at home, classroom, field)<br>- Setting (peer to peer, whole class, alone)<br>- Computer use (yes, no) |
| Tool |  | Tool type (e.g., simulation tool, modelling tool, note taking tool) |
| Activity |  | Activity type as suggested by the LAS type (e.g., build model, run experiment, define ) |
| Scaffolds |  | Scaffold that is connected to a tool or LAS that represents scaffolding mechanisms for the specific entity. |
| Resources |  | Resource that is used for an activity (e.g. video, audio, text) |

The table above (Table 1) shows the objects that are currently implemented in SCY-SE and part of the SCY language. In the study, SCY-SE was modified so that ELOs did not need to be specified with respect to format etc. and scaffolds were not available in the toolbox.

## Background: Blackboard architecture and ontology

The SCY environment relies on loosely coupled components with intelligent agent support. Agents communicate with other system components as well among themselves using a so-called *blackboard architecture*. The basic idea of the blackboard architecture is to provide interaction between components (such as agents) not by addressing other components directly, but by communicating only via writing and reading messages on the blackboard. The advantage of this paradigm is that every component is running autonomously and little knowledge about other components is necessary. This approach leads to a more robust system and provides an architecture that can be easily and flexibly extended or modified. A well-known technical approach to implementing a blackboard architecture is TupleSpaces (Gelernter, 1985). Here, a dedicated server acts as an operational centre for exchanging information. The information is stored on this server in form of tuples, which are ordered sequences of typed data. There are several implementations of TupleSpaces like JavaSpaces (Sun Microsystems) or TSpaces (IBM). We are using an implementation called SQLSpaces (Weinbrenner, Giemza, & Hoppe, 2007). SQLSpaces offer some extra features such as asynchronous call-backs, programmable expiration time of tuples, blocking and non-blocking operations, a web-based visualization, versioning and user management. Furthermore, SQLSpaces come client interfaces for several programming languages, including Java, C#, Ruby, Prolog and PHP. In the case of SCY-SE, the agent that calculates the similarity between scenarios is written in Prolog, a logic programming language.

SCY-SE also makes use of a pedagogical ontology, which captures the basic terms describing activities, tools, object types and their interrelationships. The checking mechanism, which is described later, utilizes these relations to check whether the relations between the concepts of the user are valid or not. Using the ontology for the concepts and relations results in a system, which is easily extend- and adaptable. New constraints or types of concepts can be inserted into the ontology using an editor like Protégé[3]. Using an ontology allows us not only to access concepts related to SCY, but also to specify relations between these concepts, which is useful for determining similarity between concepts.

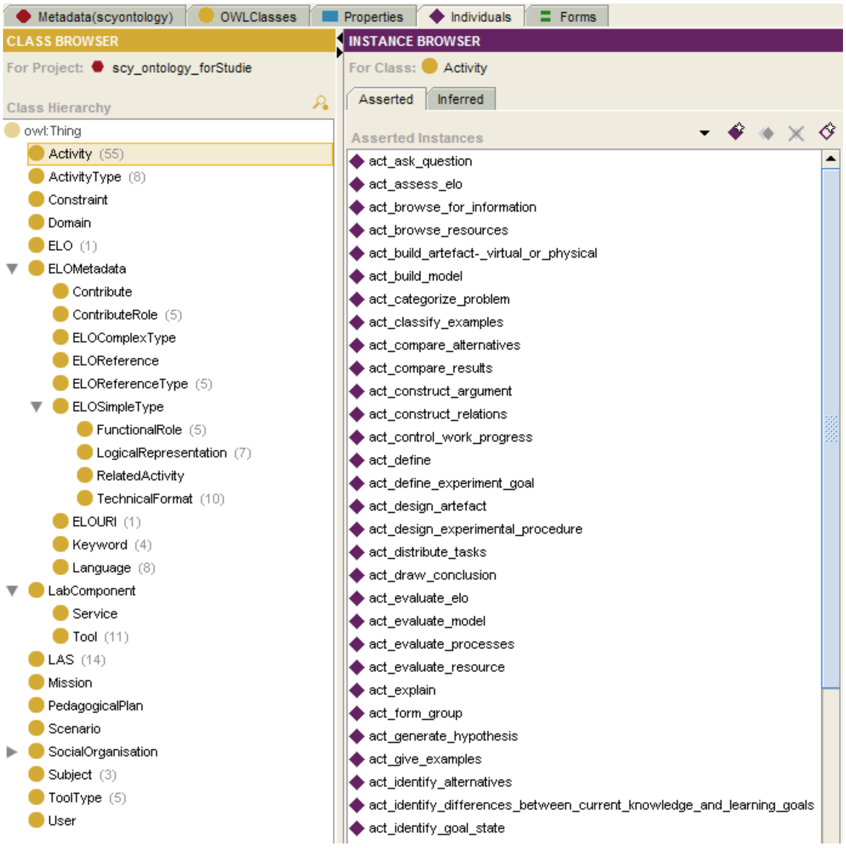



Figure 1: Protégé screenshot of the ontology and its concepts.

## Similarity algorithm

The overall process of calculating the similarity is shown in Figure 2. In the first step, SCY-SE constructs lists of the relevant entities of the scenarios to be compared. These relevant entities consist of the three components ELOs, tools and activities, which are grouped by the corresponding LASs of the two scenarios. These can again be divided into two categories: Activities and tools are atomic values, so that lists of them can be compared directly through the “weighted symmetric distance” of two sets. This distance is based on the cardinality of the symmetric difference, which is then divided by the cardinality of the union of the two sets A and B:

$$\Delta_{A,B} = \frac{\#((A \cup B) \setminus (A \cap B))}{\#(A \cup B)}$$

This distance calculation has been modified by component weights based on the ontological proximity between the single elements. The resulting similarity value ranges between 0 and 1.

ELOs, on the other hand, are not suited for a plain comparison as mentioned above. In fact, they are built up from four properties and therefore have an intrinsic complexity that requires a different approach for calculating the similarity. At first glance, it is possible to calculate the similarity of two ELOs A and B by comparing the four properties, so if two ELOs share three of their four properties, they would have a similarity of 0.75. Nevertheless, the comparison of A and C might result in an even higher similarity. So, each ELO of the first list needs to be compared with each of the second one. The outcome of this step is a matrix of similarity values. To condense this matrix to one single similarity value that expresses the similarity of these two lists, we decided to use the so-called “Hungarian method” for graph matching (Kuhn, 1955) to solve this assignment problem (to determine which ELO is identified with which other ELO).

[3] For a documentation and download location visit: http://protege.stanford.edu/doc/users.html (last visited in October 2009)

As shown in Figure 2, the calculation inside the Prolog agent is done in several steps. After the lists have been written to the SQLSpaces server, a "start comparing" tuple is written to inform the agent about a new query. Then the agent fetches the lists and calculates the weighted symmetric distance for each of the lists. At that point the matrix for the ELOs is already calculated by using the "Hungarian method". The next step uses the merged matrix of these three matrices. The Hungarian method is then used again to determine the best mapping between the LASs of the two scenarios with respect to the similarity encoded in the matrix. The method may take several iterations, but will eventually terminate and return the optimal solution. From the resulting mapping, a single value can be easily calculated by normalizing the selected matrix entries. The returned "similarity value" (Sim(A,B)) is transformed in such a way that Δ=0 corresponds to Sim=100% and Δ=1 corresponds to Sim=0%.

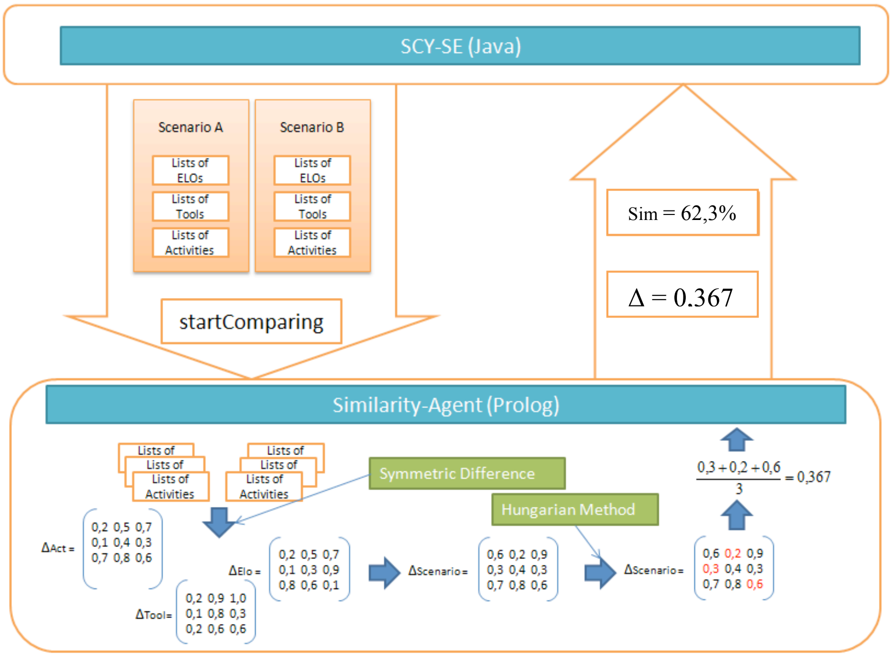


Figure 2: The overall process of the similarity measurement.

For the study, not all components of the similarity measure had an effect on the result because of the limited version of SCY-SE that participants were using. Since the limited version did not include options to specify ELOs with respect to technical format etc., all ELOs were similar. Therefore, the similarity measurement was based on other factors (e.g., number of LASs, number and type of tools and activities).

## Checking

SCY-SE allows for the automatic checking against several types of constraints and constraint violations (such as incompatibilities of object types) in given designs, some of which are ontology based. The more obvious ones apply to the way nodes are connected. To distinguish between ELOs that are emerging from a LAS and ELOs that work as starting material for a LAS, SCY uses the notion of input and output ELOs. An output ELO needs some activity that produces it and an input ELO needs some activity that consumes it. Moreover, ELOs can have more or less precisely defined types. These types also have a "compatibility" relation to tools (that support activities) based on the tool's potential output data formats. This compatibility relation is stored in the ontology. Violations of any rule are displayed next to each incorrect node with short information about the problem and possibly a hint for resolving this violation.

## Research goals and questions

The main goal of this study was to test the validity of the similarity algorithm. We prepared three textual descriptions of scenarios consisting of inquiry learning activities. Based on these textual scenarios, students created graphical scenarios using SCY-SE. In addition, expert solutions were created beforehand by researchers, to which the student scenarios were compared. To distinguish between student scenarios and expert solutions, scenarios created by students use a lower case letter (a, b, c) and expert solutions use an uppercase letter (A, B, C). The following questions were of interest:

1. a) Are SCY-SE scenarios created based on the textual scenario a, more similar to a corresponding expert solution A than to a non-corresponding expert solution B?
   Hypotheses: Sim(a, A) > Sim(a, B); Sim(a, A) > Sim(a, C); Sim(b,B) > Sim(b,A) etc.

b) Are the differences induced by the given textual scenarios still discernable in the graphical scenarios produced by the same subject?
Hypotheses: Sim(a, b) < Sim(a, A); Sim(a, c) < Sim(a, A); Sim(b, c) < Sim(b, B) etc.

Additionally, we were interested in the influence of checking and ensuing feedback:

2. a) Does the checking feature help to guide users during the process of creating a complete SCY-SE scenario?
   b) Does using the checking feature increase the similarity to the corresponding expert solution?

## Design of the Study

### Participants, procedure and instrument

Twenty-five students of media-technology or teaching were participating in the study, taking place at a university lab in Germany. Participants listened to a half hour introduction to the SCY-SE approach and to the language including a demo of how to use SCY-SE. After the introduction, everyone received a small manual and the task description. Every student worked individually on one computer. The task was to translate three textual scenarios into graphical scenarios using SCY-SE. Participants had to decide which LASs, activities, tools and ELOs need to be selected to represent the textual scenarios graphically. When a student was finished with the last scenario, she was requested to save the *SCY-SE scenario c* under a different name, to use the checking button, and to revise *SCY-SE scenario c-checked* accordingly. The participation time was two hours. A questionnaire including demographic questions and single-item questions with respect to perceived difficulty was administered right after the study.

### Material

Every textual scenario consisted of four paragraphs. Each paragraph represented one LAS. All textual scenarios described activities following an inquiry approach within the subject matter of photosynthesis. The textual scenarios differed from each other with respect to activities, tools and ELOs. The following paragraph is an example of how activities were described in the textual scenario:

> "Students work in groups to do their experiments using SCY Simulator. They define their experiment goals in SCY Simulator. Afterwards the students start doing their experiment runs. Then, the students compare the results with their hypothesis stated earlier. They make notes, which they upload together with their experiment data."

The translation of this paragraph into a SCY-SE scenario may look like this:

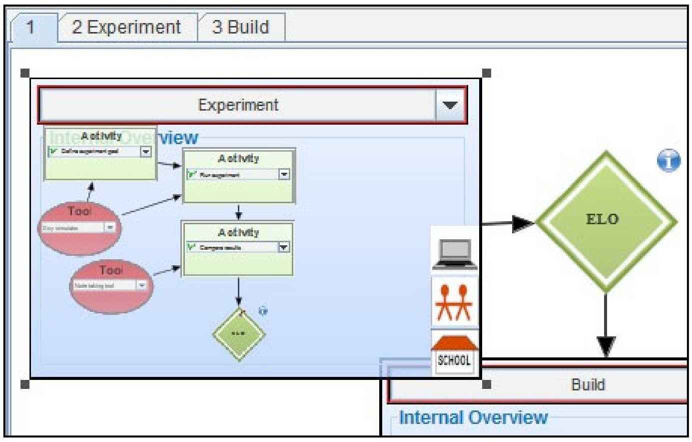



Figure 3. SCY-SE Scenario view

### Source of Analysis

Hundred SCY-SE scenario files (four – three unchecked and one checked- from each participant) were collected and analyzed. Similarities were calculated. For answering research question 1a, SCY-SE scenarios were compared with corresponding and non-corresponding SCY-SE scenario expert solutions. Concerning question 1b, SCY-SE scenarios were compared with SCY-SE scenarios by the same student. Answering research question

2a required logging the number of errors detected by the checking mechanism and question 2b required to compare SCY-SE scenario c with SCY-SE scenario c-checked.

## Results

Due to low sample size, descriptive statistics were favoured and non-parametric methods were used. From the questionnaire, we found that most participants (72%) reported that textual scenarios were of similar difficulty level. Some participants (24%) identified one of the three scenarios as most difficult. Most participants (76%) reported that it was most difficult to select the tools (besides LASs and Activities).

First, we compared every student's scenario a, b, c with the corresponding expert solution and with the non-corresponding expert solutions. As expected, if student scenarios were compared with the corresponding expert solution, the percentage with respect to similarity was higher (M=81, SD=07) than if student scenarios were compared with the non-corresponding expert solutions (M=55, SD=04). Wilcoxon's Signed-Rank test showed that this difference (Table 2 and Table 3) was statistically significant (z=-4.37, p=.000).

Table 2. Descriptives for similarity between student scenario and corresponding expert solution in %

| student scenario \| expert solution | N | Min | Max | Mean | Std. Dev. |
|---|---|---|---|---|---|
| a \| A | 25 | 25 | 85 | 74 | 12 |
| b \| B | 25 | 62 | 99 | 83 | 09 |
| c \| C | 25 | 62 | 96 | 86 | 08 |

Table 3. Descriptives for similarity between student scenario and non-corresponding expert solution in %

| student scenario \| expert solution | N | Min | Max | Mean | Std. Dev. |
|---|---|---|---|---|---|
| a \| C | 25 | 24 | 68 | 54 | 10 |
| a \| B | 25 | 19 | 69 | 55 | 10 |
| b \| C | 25 | 40 | 68 | 54 | 06 |
| b \| A | 25 | 44 | 69 | 56 | 05 |
| c \| A | 25 | 47 | 55 | 58 | 05 |
| c \| B | 25 | 41 | 67 | 48 | 04 |

Second, we compared student scenarios with student scenarios (Table 4) done by the same student (Sim(a,b); Sim(a,c), Sim(b,c)). As expected, similarity was low (M=53, SD=08) and comparative to similarity between expert solutions A, B, C (M=58, SD=49).

Table 4. Descriptives for similarity between (non-corresponding) scenarios for every student

| student scenario \| student scenario | N | Min | Max | Mean | Std. Dev. |
|---|---|---|---|---|---|
| a \| b | 25 | 14 | 68 | 55 | 10 |
| a \| c | 25 | 21 | 75 | 53 | 11 |
| b \| c | 25 | 39 | 65 | 51 | 06 |
| Valid N (listwise) | 25 | | | | |

### Checking

Results show, that the checking feature was necessary for most participants to create complete and correct SCY-SE scenarios. On average, participants made 4.7 (SD=3.3) errors that were detected by the checking mechanism. In order to find out whether the checking feature would increase similarity between student scenario and expert solution, we compared the last scenario (*SCY-SE scenario c*) that was developed with the corresponding expert solution (*SCY-SE Scenario expert solution C*). We found no significant difference with respect to similarity comparing unchecked and checked scenarios. On average, *SCY-SE scenario c* was 86% similar to the *SCY-SE scenario expert solution C*. After using the checking mechanism, average similarity was still 86%. In order to find out why the checking feature did not yield higher similarity, an in depth analysis with respect to the types of errors was done. The errors that were displayed while using the checking feature were mostly related to incorrect usage of links between objects or missing links. For example, many participants missed making links between tools and activities. Since the similarity algorithm did not take into account the number of links, the checking had no positive effect on the similarity measure.

## Conclusion and Implications

The results demonstrate that the similarity measure captures well the plausible similarity between graphical scenario designs originating from the same textual source across subjects (as “modellers”). We could also show that dissimilarities of scenarios developed by the same person but based on dissimilar textual scenarios were detected (and therefore discriminated). It was somewhat surprising that scenarios that were based on non-corresponding textual descriptions still resulted in a similarity of over 50%. One reason was that all three textual scenarios (which were the basis for the graphical representations in SCY-SE) had approximately the same length and the same number of LASs. Together with certain tool similarities suggested by the ontology this resulted in a quite high baseline. Yet, as the similarity of scenarios based on corresponding textual scenarios was over 80% (which results in a statistically significant difference), we were able to reliably discriminate similar scenarios from dissimilar scenarios. Another finding was that guiding a practitioner during the creation of a scenario using SCY-SE is beneficial. Using a checking mechanism helps to reduce the number of errors and to use the SCY graphical language consistently.

## References

Anjewierden, A., Chen, W., Wichmann, A., & Borkulo van, S. P. (2009). Process and product analysis to initiate peer-to-peer collaboration on system dynamics modelling. In J. Bourdeau and R. Mizoguchi (Eds.), *Intelligent and Innovative Support for Collaborative Learning Activities (WIISCOLA).* Workshop conducted at the Computer-Supported Collaborative Learning Conference CSCL 2009, Rhodes, Greece.

Dalziel, J. R. (2003, August). Implementing learning design: the learning activity management system (LAMS). *Paper presented at the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*. Adelaide, Australia.

De Jong, T., Joolingen van, W. R., Weinberger, A., & the SCY team (2009, June). Learning by design. An example from the SCY-project. *Paper presented at the CSCL 2009*, Rhodes, Greece.

Francq, P., & Delchambre, A. (2005). Using documents assessments to build communities of interest. *Proceedings of the IEEE/IPSJ International Symposium on Applications and the Internet* (SAINT 2005) (pp. 237-333). Los Alamitos, USA: IEEE Computer Society.

Gelernter, D. (1985). Generative communication in Linda. *ACM Transactions on Programming Languages and Systems, 7*(1), 80-112.

Hoppe, H. U., & Gaßner, K. (2002). Integrating collaborative concept mapping tools with group memory and retrieval functions. In G. Stahl (Ed.), *International Conference on Computer Supported Collaborative Learning: Foundations for a CSCL Community* (pp. 716–725). Mahwah, NJ: LEA.

Hoppe, H. U., Pinkwart, N., Oelinger, M., Zeini, S., Verdejo, F., Barros, B., Mayorga, J. I. (2005). Building bridges within learning communities through ontologies and “thematic objects”. In *Proceedings of* Computer-Supported Collaborative Learning Conference CSCL 2005 (pp. 211-220). Mahwah, NJ: LEA.

Konstan, J.A. & Riedl, J. (2002). Collaborative Filtering: Supporting social navigation in large, crowded infospaces. In Höök, K., Munro, A., & Benyon, D. (Eds.). *Designing Information Spaces: The Social Navigation Approach* (pp. 43-82). Berlin, Germany: Springer.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2, 83-97.

Miao, Y., Harrer, A., Hoeksema, K., & Hoppe, H. U. (2007). Modelling CSCL scripts – a reflection on learning design approaches. In F. Fischer et al. (Eds.), *Scripting Computer-Supported Collaborative Learning* (pp. 116-134). Berlin, Germany: Springer.

Ronen, M., Kohen-Vacs, D., & Raz-Fogel, N. (2006). Structuring, sharing and reusing asynchronous collaborative pedagogy. In Y. Kafai, W. Sandoval, A. Noel Enyedy & F. Herrera (Eds.), *International Conference of the Learning Sciences* (Vol. 1, pp. 955-605). Bloomington, Indiana: LEA.

Weinbrenner, S., Giemza, A. & Hoppe, H. U (2007). Engineering heterogeneous distributed learning environ-ments using TupleSpaces as an architectural platform. *Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies* (pp. 434-436). Los Alamitos, USA: IEEE Computer Society.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.

# Preparing for the Long Tail of Teaching and Learning Tools

Charles Severance and Stephanie D. Teasley, School of Information, University of Michigan,
1075 Beal Ave, Ann Arbor, MI, 48109-2112
Email: [csev, steasley]@umich.edu

**Abstract:** In this paper we apply the concept of "the long tail" (Anderson, 2006) to teaching and learning tools to discuss how the limitations of current Learning Management Systems (LMS) can be overcome to allow instructors to customize the technology they use to support their own classroom practices. Learning tools in the long tail are those that are widely used by a subset of instructors - tools specific to large courses or tools specific to a particular field, and tools that are only used in a few courses or a single course. Using several examples from courses taught on our campus, we show how to put extensibility in the hands of the instructors to create knowledge-age learning technologies that are customizable, interactive and controlled by users.

## Introduction

Increasingly, learning management systems (LMS) such as Blackboard or Sakai are seen as one of the most essential Enterprise Services in education. A recent survey of 115 American universities has shown that 89% of students reported that they had taken a course that used a LMS (Smith, Salaway, & Caruso, 2009). In addition, a 2009 survey of US school district administrators estimated that more than a million k-12 students took online courses in the school year 2007-2008 (Picciano & Seaman, 2009). That these systems are basic infrastructure for learning in higher education is already a fact; that they are as common in K-12 education may also soon be true (see Means, Toyama, Murphy, Bakai & Jones, 2009). Yet what do we know about using LMS well for teaching and learning? How can we help teachers to incorporate the promise of Web 2.0 technologies into their classrooms? In this paper, we examine the trends in the evolution of learning management systems and how those systems are currently being used in higher education. We then propose how learning management systems must change by leveraging the "the long tail" (Anderson, 2006) of teaching and learning tools. These suggestions apply to higher education specifically, but also provide guidelines for development and use that may ease the transition as LMS use permeates K-12 education.

## Background

Few other campus enterprise systems have the requirements of 24x7 availability, with the ability to scale to support over 10,000 simultaneous users during peak loads. On many campuses, learning management systems must run for months without allowing for an outage to perform major software upgrades. These requirements lead to a very careful and conservative approach to upgrading or changing the LMS software in the middle of a semester. This trend is coupled with increasing penetration of the LMS system as measured by the percentage of students and faculty who are using the LMS (Smith, Salaway, & Caruso, 2009). On our own campus, the annual IT survey shows that 99% of students and 81% of faculty have used our LMS for at least one course in the past year (Lonn & Teasley, 2009).

Because LMS use has become so pervasive in higher education, we have an opportunity to analyze how the average instructor uses these systems across many different subjects. Most analyses of LMS use (Hanson & Robson, 2004; West, Waddoups, & Graham, 2007) point to a distribution that follows the "long tail" typically found in analyses of most online systems (Anderson, 2004). Specifically, the long tail refers to the statistical phenomenon of a power law or Pareto distribution where few items comprise the most use but there is a long tail of many items used with a much lower frequency. This distribution is clearly seen with the use of tools available within LMS, where a few tools are heavily used then usage drops off dramatically after five or six core tools. Overall, document management and broadcast-oriented communication tools (Content Sharing, Assignments, Announcements, Schedule, and Syllabus) comprise 95% of all user actions (Lonn & Teasley, 2009; Hanson & Robson, 2004). By contrast, the tools that are more interactive (Chat, Discussion, and Wiki) are not used as much. While research coming from the Learning Sciences would have much add to the current literature addressing the relative value of teaching with one tool or another, in this paper we leave this to others and focus here on the fact of current LMS use and how to empower instructors improve their own use of these systems.

The two trends of "LMS as critical infrastructure" and "only a few of the tools are heavily used" leads to the inevitable conclusion that LMS development efforts need to focus on improving the core tools which make up the LMS and spend less effort on the long tail of tools. If the trend was extended to infinity, LMS systems of the future might have exactly seven tools which are never changed or upgraded. While this will insure that the core infrastructure is solid, consistent, and reliable, it will have a tremendous negative impact on the ability for instructors and learners to innovate and find new ways to use technology in education. One possible path forward is that learning management systems will go "underground" - where in order to experiment with innovative ideas, savvy faculty host their own learning management systems under their desks or perhaps run software on their own ISP account. The "Edupunk" movement (e.g., Kuntz, 2008; Young, 2008) expresses this sentiment in a call-to-arms to reject commercial LMS products. This approach does allow instructors to be innovative, but it adds the burden of maintaining a production infrastructure and saps precious energy away from their teaching efforts. Another extreme reaction to the perceived limitations of current LMS is a call to "teach naked" and reject the use of technology in the classroom altogether (Young, 2009). This, however, seems like a “baby with the bathwater” solution that is not likely to be realistic for today’s students who are considered to be the “net generation” and “tech-savvy Millennials” (Junco & Mastrodicasa, 2007).

Rather that going Edupunk or even teaching naked, we believe the solution to this problem is to add features to LMS systems that allow the core functionality to focus on scalability and stability while allowing innovation at the edges by encouraging more use in the long tail. The key to this approach is that we need to add features to LMS systems so that they can be extended without adding a new feature on to the LMS servers or needing to upgrade the LMS to a new version (Severance, Hardin, & Whyte, 2008). The extensibility needs to be placed in the hands of the instructors rather than only in the hands of the LMS system administrators. This DIY (Do It Yourself) attitude reflects the growing capacity of Web 2.0 applications to put users in control of the content and distribution of materials. In popular culture this DIY capability can be seen in zines, self-publishing, and music re-mixes. We believe this approach can be extended to educational tools as well and fulfill Collins & Halverson’s (2009) call for knowledge-age learning technologies to be customizable, interactive and controlled by users. Only then can we meet both the needs of enterprise production and innovative approaches to teaching and have the best of both worlds. The average instructor who only uses 5-6 core tools has access to a scalable and stable toolset, while the instructor with a new idea is allowed to bring that idea into their class in a few days or weeks of effort - all without destabilizing the LMS production system.

## Teaching Tools in the Long Tail

We see learning tools falling into three basic categories: (1) the core 5-10 tools used by nearly every teacher, (2) a set of tools that are widely used by some subset of the teachers - perhaps tools specific to large courses or tools specific to a particular fields like mathematics, and (3) tools that are only used by a few courses or even a tool purpose built for a single course. As we look at the nature of the tools in (1) and compare them to the tools in (3), we are likely to see a transition from tools that "manage" the learning process towards tools that support the learning process. The tools in category (2) are likely a mix of management and learning. This leads to a "long tail" effect where the more learning-oriented tools are in the long-tail. While each individual tool may have a very small "market share" when aggregated together, these long-tail tools may well represent a majority of the overall usage.

The nature of the content-oriented tools (category 1) that are used universally is that their features are likely to be useful to every single instructor, regardless of context or discipline. Hence Category 1 tools comprise the bulk of the distribution of use curve. By contrast, the tools in the second category tend to end up appealing to a smaller but identifiable population of instructors. For example, the CAPA testing system uses LaTeX as its question authoring language and as such naturally appeals fields such as mathematics, chemistry, and physics where most of the instructors with a Ph.D. in those fields learned LaTeX to write and publish papers. Furthermore, CAPA provides a very rich (albeit complex) mechanism for generating many equivalent variations of a problem by altering numeric values randomly. This functionality is very useful for courses where many of the problem sets assigned to students involve numeric calculations. While the CAPA system is very popular for use in first and second year physics, math, and chemistry classes with high-enrollment numbers, it is simply too difficult to learn to ever become widely used for fields like literature or the humanities. This naturally limits the *overall* number of courses and faculty who will use a CAPA-based system to a small fraction of the market – perhaps less than 2-3% of the overall courses taught. However, for those courses, CAPA is nearly the *perfect* solution particularly when coupled with the ability to collectively build large question pools across institutions and with some publishers providing CAPA question banks with physics and chemistry textbooks. Since CAPA reflects such a small market share overall, the testing systems provided in mainstream LMS products do not include CAPA-like features and so if you teach a course that needs CAPA – pretty much your *only* choice is CAPA. For tools in the third category, the potential market share is smaller yet and these tools may have very individualized use that can not necessarily be generalized across disciplines or teaching contexts.

In what follows below, we discuss several examples of category 2 & 3 tools in the long tail and provide detail about the ways in which these tools extend instructors’ use of the standard LMS toolset to meet their unique needs. These examples reflect current teaching practice at the University of Michigan where the enterprise LMS is based on the Sakai open-source LMS.

**Student Assessment Management System (SAMS)**

In our College of Literature, Science, and Arts (LSA), instructors have access to a CAPA-based system called SAMS (Student Assessment Management System) which is heavily used by the physics, mathematics, and chemistry departments. A requirement unique to SAMS is the need to do extensive data mining across the multiple sections of the same course. Since there are so many sections taught by graduate student instructors in introductory-level courses, SAMS must be able to provide regular reports to course coordinators so that problems encountered by individual student-instructors can be diagnosed and addressed as quickly as possible. In addition, error patterns in problem sets seen across sections provides the main instructor with feedback about which concepts and/or formulas need further elaboration in lecture or additional time in section. For these reasons, SAMS is considered to be a powerful tool in achieving consistently high quality in the teaching of these large-enrollment courses.

Despite its important role in the largest academic unit on campus, SAMS is not in the standard toolset provided by the LMS. Since SAMS is written in PERL and the underlying architecture of the LMS (Sakai) is written in Java and because the requirements for SAMS are so complex (e.g., including rules about to who can see which data and reports), it was never practical to re-write SAMS inside of Sakai. For many years students in classes that used SAMS visited two separate sites for their courses: one course site in the LMS and one course site in SAMS. This was confusing and inconvenient for students *and* instructors as the SAMS site had its own navigation, login process, and user interface conventions. After an early version of IMS Learning Tools Interoperability was installed in Sakai, we were able to integrate SAMS into Sakai to share identity and roster information with SAMS without any user intervention. We even created a virtual tool in Sakai that made it look like we have built SAMS into Sakai. Instructors can now simply add the SAMS tool to their course site like any built-in Sakai tool. Figure 1 displays the user's view of SAMS inside of a Sakai course site.

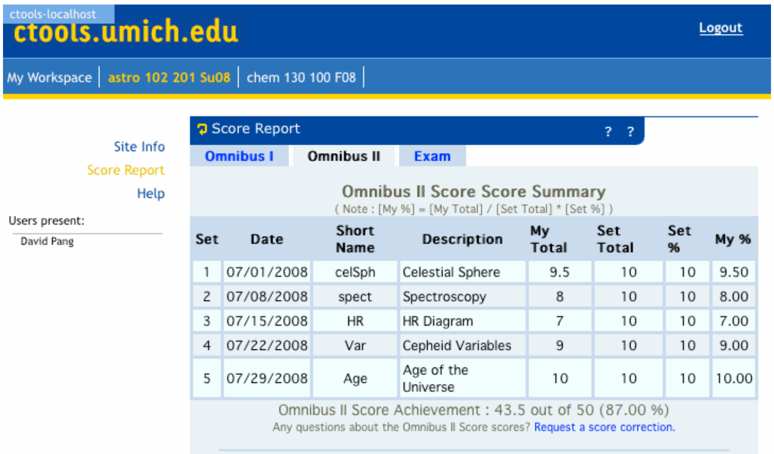



Figure 1. SAMS Running Within Sakai

Effectively the user experience for both the instructors and students is as if the SAMS tool had been ported into Sakai. There is no need to rewrite any software; we only had to add some integration in SAMS to receive and process the IMS Learning Tools Interoperability launch requests from Sakai. This approach also allows the College of LSA to maintain their strategic access to their data, and to independently upgrade and improve SAMS to meet their needs on their own schedule, unencumbered by the Sakai development or production priorities.

This is an excellent example of how we can develop category 2 tools to meet both the enterprise-wide needs in teaching and learning as well as the school-level needs for teaching and learning. The approach allows reuse of the common capabilities of the enterprise systems while allowing schools or departments to address their own unique needs in focused areas of teaching and learning. An enterprise LMS does not have to be a win-lose proposition across campus.

**LectureTools**

The LectureTools project provides free tools that support interactivity and enhanced modes of learning during lectures. Like CAPA, LectureTools is most useful for medium to large lecture courses where the teaching staff wants to use support interactions between the instructor and student, and between students as part of the lecture experience. Again we see a situation where the overall population of instructors for whom LectureTools is useful is a fraction of the entire set of courses that are taught at the university. And, here again, the functionality provided by Lecture Tools is not likely to be included into the core functionality of most LMS systems.

Like the SAMS project, we developed a similar virtual tool approach to integrate LectureTools into Sakai. Instructors can add the LectureTools tool to their course site like any other tool built into Sakai. Sakai uses IMS Learning Tools Interoperability to launch and provision course sites in LectureTools, giving students and instructors a seamless user experience from a single course site. Unlike SAMS, the LectureTools service is available to instructors at any university in the US and Canada. These additional schools may or may not use Sakai as their LMS. As IMS Learning Tools Interoperability support is added to all LMS, any school can integrate LectureTools into their campus-wise enterprise LMS systems.

We are currently in the middle of a project integrating LectureTools into Blackboard LMS running at a community college and one commuter campus of the large research university as shown in Figure 2. This project will not only demonstrate the ability of LectureTools to run regardless of which enterprise LMS is in use, but provide a model for allowing cross-campus collaboration in teaching specific courses.

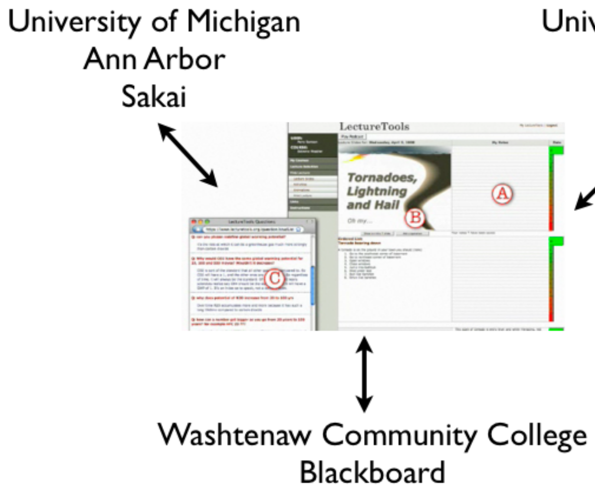



Figure 2. Using LectureTools in Sakai and Blackboard

We are using an open-source Blackboard Building Block that supports IMS Learning Tools Interoperability developed by Stephen Vickers of Edinburgh University [www.spvsoftwareproducts.com]. Once the IMS Tools Interoperability integration is completed, the same tool can be used across these three institutions with each set of users experiencing the tool seamlessly integrated into their local LMS user interface. As this pattern is extended, it allows a cross-institutional community to develop where the common thread is the use of the LectureTools platform to augment their lecture experiences. By combining small pools of interest across many campuses, is it possible to end up with a much larger overall demand for a tool or capability. By reducing the integration costs to nearly zero using IMS Learning Tools Interoperability, we increase the likelihood that these cross-institutional communities will form around particular pedagogy or domain specific tools.

In summary, the middle category of tools, Category 2, are those tools that appeal to some subset of the overall teaching space and are very valuable to that those teachers and learners. By allowing tools to be scoped at a college or department level or perhaps by bringing a cross-institutional community of interest together, we can match the tool with its level of demand. While the core tools are very focused on the management of learning, the tools in the middle category are generally some combination of "learning management" and content or context specific learning. That is, these more narrow tools will often focus on supporting a particular teaching pedagogy or objective rather than simply moving content around and facilitating students' access to that content.

## Wisdom of Crowds

In addition to Category 2 tools that have broad use with a small market segment, there are also tools that only appeal to a very tiny population – perhaps as small as a single instructor (Category 3). The exemplar for this category of tools comes from the book "Wisdom of Crowds" by James Surowiecki (2005). Surowiecki's book provides examples of how groups of people can have collective intelligence that surpasses the intelligence of any of its individual members. The author uses examples from social science, economics, and game theory to provide a basis to explain the mechanisms that make crowds wise. Often these concepts are explained in the form of a multi-player game where students play a game and then afterwards the class analyzes player behavior to illustrate the point of the exercise.

Often when teachers use the book "Wisdom of Crowds" they play the games with small scraps of paper and a designated student who "runs" the games. However it is also possible to write computer software that simulates the game and enforces its rules. When computer software is used to run the game, the educational advantage is we can retain the history of player interactions to facilitate a deeper insight into why a player made a move at a particular moment in time. For example, the simplest of the games proposed by Surowieki has a group guess a numeric value such as the number of jellybeans in a jar. First people independently guess the value and then an average the values is calculated.

This game can be played much more effectively on a computer using student laptops or PDA's rather than averaging numbers on slips of paper. Using technology, students experience first hand how their own guesses may be less accurate that the group mean and the visible display of individual guesses shows more clearly how the collective arrives at the correct answer. To implement the software for the game, we built a simple application that consisted of 118 lines of Python code hosted in the Google Application Engine cloud environment. The tool handled the IMS Basic LTI protocol and implemented the rules of the guessing game. The instructor could reset the game or view the results – the students could simply make a guess. The tool was then integrated into Sakai using IMS LTI and made available in the course site as shown in Figure 3.

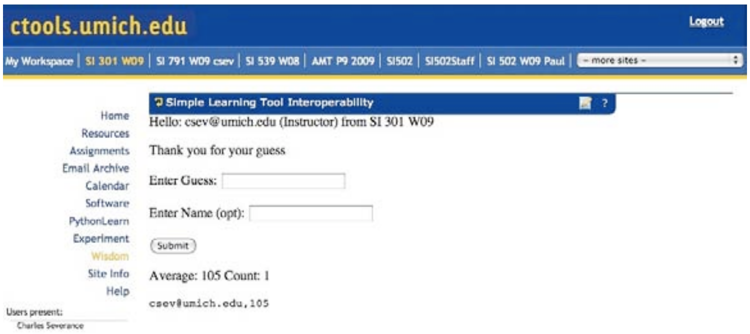



Figure 3. The Number Guessing Application Running in Sakai

The number guessing game was written in about two hours and used in lecture on the same day that it was written. After the game was used for one lecture, there were a few bugs that were found and fixed for use in later lectures. A tech-savvy instructor did the entire process with no impact on, nor involvement of, the enterprise Learning Management System. And since the tool was hosted for free on the Google Application Engine, the instructor did not even have to worry about the infrastructure needed to run the tool.

Another game was written to demonstrate the "Free Rider" problem which occurs when groups are sharing the costs of some shared common good and how people balance the overall group benefit against their own short-term potential for gain. The Free Rider Application had several features that made it very effective for in-class use. First, since the game enforced the rules, it was not necessary to teach anyone how to "run" the game. Also, the game picked five students to play the game automatically. Once the players were selected and the game started, the other students were given a display that updated dynamically as the game was played. So students could learn by playing the game and when they were not playing, they could watch as game masters. The students who were watching could see when players changed strategies and could see the game develop and see which strategies led to the largest payoff.

The games are very simple and easy to write – since they are embedded in a rich LMS, the tools only have to solve the very simple problem related to the lesson at hand. Once these tools are written and put up on Google Application Engine, they could be used by any instructor using the "Wisdom of Crowds" book in their classroom by simply exchanging the IMS Learning Tools Interoperability URL, Key, and Secret.

The number of teachers using "Wisdom of Crowds" in their classroom at any given moment or during any given semester is very small. But at the same time, the effort to develop and the tools is also very small. And the effort involved is small enough that a single teacher might do it simply for his or her own use. Following the example of free applications available in an "apps store" this instructor might also post it on a public site for any other instructor using Surowiecki’s book in their course. This example is toward the far end of the long tail of teaching applications. However, even if it only affects 25 courses across the country in any semester, these tools can be designed to be really helpful for helping students to understand more deeply this material. One could imagine a future where books like "Wisdom of Crowds" might come with already-built games developed and provided by the author or publishers. These games would be ready to plug into the local enterprise LMS using IMS Learning Tools Interoperability.

## Required LMS Features to Enable the Long Tail

If we are to address the need to build and use the long tail of learning tools, we must reduce the barriers to plugging new tools into Learning Management Systems. Opening up these system to outside applications ultimately puts the ability to "add a tool" in the hands of the instructors and allows them to add the new tools in a few clicks and with no intervention on the part of the technical support staff. Sakai is generally designed to give instructors a great deal of control of course content to the instructors. A Basic LTI tool has been developed for Sakai that allows the instructor to easily integrate externally provided tools into Sakai. The primary information needed for to integrate a tool using Basic LTI is a URL, Key, and Secret as shown in Figure 4.

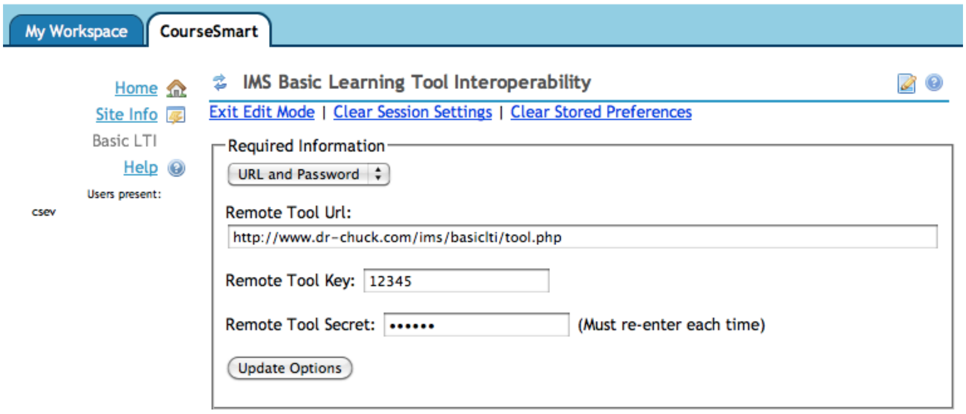



Figure 4. Setting the URL, Key, and Secret in the Sakai Basic LTI Tool

Since the IMS Basic LTI tool will send roster information to the externally provided tool, it is important to make sure that the instructor is aware that this is happening and approves the release of any identifying information using the configuration options shown in Figure 5.

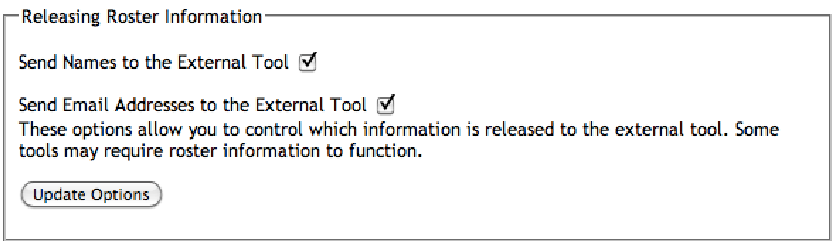



Figure 5. Privacy Controls in the Sakai Basic LTI Tool

The IMS Basic LTI specification makes any data that contains identifying information optional. The default in Sakai is not to send any identifying information so the teacher must explicitly agree to send the identifying information to the external tool.

The developers of each LMS will make their own choices about which aspects of LTI are placed in the hands of instructors and which aspects of configuration are only available to system administrators or technical support staff. The Sakai tool allows local customization of the configuration process for LTI, giving system administrators fine-grained access control over which features and capabilities are made available to the Instructors. This allows each institution to progress toward the model of many tools in the long tail at the pace that is comfortable and sustainable for their organization.

## Conclusion

We present the case for adding more flexibility to Learning Management Systems and putting that flexibility in the hands of instructors. By making it possible to easily integrate more narrow and learning-centered tools into the LMS without requiring a change in production software or server reboot, we make it far more practical for teachers and students to experiment with new tools and to find the right set of tools for their particular course, supporting a move from accidental to intentional pedagogy (McGee, Carmean & Jafari, 2005). Once the barriers are removed from within the LMS, a market for these externally hosted tools can develop– particularly in the "middle tail" category where tools have broad applications within a narrow segment of the population. We would hope that many commercial and free tools would be developed and made easily available – resulting in many innovative experiments that can lead to a greatly improved learning experience for students of any age.

Once the barriers for implementation are reduced even further, we envision that tools will be written by teachers or students to solve very focused learning needs. As LMS evolve and interoperability standards improve, many of these tools will be very simple to develop and use because they will be placed in the rich context of a mainstream LMS.

Perhaps the most exciting aspect of enabling teachers to build, exchange, and use thousands or even hundreds of thousands of new tools is how we enable the exploring of an increasingly wide range of new ways to teach. In addition, by opening the enterprise LMSs to virtually unlimited expansion, we have a place to explore emerging approaches such as social learning and the increased use and remixing of content from Open Educational Resources in new and novel ways. In a sense, while we can see an immediate exciting future that this approach enables, the truly exciting innovations are those that we can't even imagine because we are locked into the content-delivery patterns of the current crop of enterprise LMS. Finally, by opening up these opportunities to instructors we simultaneously open them up for students to build, organize, and use tools for their own collaboration and learning purposes.

## References

Anderson, C. (2004). "The Long Tail" *Wired*, *12* (10). http://www.wired.com/wired/archive/12.10/tail.html.

Anderson, C. (2006). *The Long Tail: Why the Future of Business Is Selling Less of More.* NY: Hyperion.

Collins, A. & Halverson, R. (2009). *Rethinking Education in the Age of Technology: Digital Revolution and Schooling in America*. New York, NY: Teachers College Press.

Hanson, P., & Robson, R. (2004). Evaluating course management technology: A pilot study. *Educause Center for Applied Research, Research Bulletin*, (*24*), Boulder, CO: EDUCAUSE. http://www.educause.edu/library/ERB0424

Junco, R. & Mastrodicasa, J. M. (2007). Connecting to the Net.Generation: What higher education professionals need to know about today's students.

Kuntz, T. (2008). The Buzz for 'Edupunk', New York *Times*. http://ideas.blogs.nytimes.com/2008/10/17/the-buzz-for-edupunk/

Lonn, S. & Teasley, S. D. (2009). Saving time or innovating practice: Investigating perceptions and uses of Learning Management Systems. *Computers & Education*, *53*, 686-694.

Means, B., Toyama, Y, Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*. Report for the US Department of Education. http://www.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

McGee, P. Carmean, C., & Jafari, A. (2005). *Management systems for learning: Moving beyond accidental pedagogy*. Hersey, PA: Information Science Publishing.

Picciano, A. G., & Seaman, J. (2007). K-12 online learning: A survey of U.S. school district administrators. Boston, MA: Sloan Consortium. http://www.sloan-c.org/publications/survey/K-12_06.asp

Severance, C., Hardin, J., & Whyte, J. (2008). The Coming Functionality Mashup in Personal Learning Environments, *Interactive Learning Environments*, *16* (1), p. .

Smith, S. D., Salaway, G., & Caruso, J. B. (2009). *The ECAR Study of Undergraduate Students and Information Technology, 2009*. Boulder, CO: EDUCAUSE. http://www.educause.edu/Resources/TheECARStudyofUndergraduateStu/187226

Surowiecki, J. (2005). *The Wisdom of Crowds*. New York, NY: Anchor Books.

West, R. E., Waddoups, G., & Graham, C. R. (2007). Understanding the experiences of instructors as they adopt a course management system. *Educational Technology Research and Development*, *55* (1), 1-26.

Young, J. R. (May 30, 2008). Frustrated With Corporate Course-Management Systems, Some Professors Go 'Edupunk'", Chronicle of Higher Education. http://chronicle.com/wiredcampus/article/3045/frustrated-with-corporate-course-management-systems-some-professors-go-edupunk

Young, J. R. (2009), "When computers leave classrooms, so does boredom." *Chronicle of Higher Education.* http://chronicle.com/article/Teach-Naked-Effort-Strips/47398/

## Acknowledgments

Thank you to Perry Samson, Bret Squire, and Francisco Roque for their assistance in this work.

# Students' Use of Multiple Strategies for Spatial Problem Solving

Mike Stieff, Minjung Ryu, Bonnie Dixon
University of Maryland-College Park, College Park, MD
mstieff@umd.edu, mryu@umd.edu, bdixon1@umd.edu

**Abstract:** In scientific problem solving, spatial thinking is critical for reasoning about spatial relationships in three-dimensions and representing spatial information in diagrams. Despite the importance of spatial thinking, little is known about the underlying cognitive components of spatial thinking and the strategies that students employ to solve spatial problems. Namely, it is unclear whether students employ imagistic reasoning strategies while engaged in spatial thinking. In the present study, we investigate which strategies students use to solve spatial chemistry problems and the relationships between strategy choice, achievement, spatial ability and sex. The results indicate that students employ multiple strategies that include the use of diagrams and heuristics rather than merely relying on imagistic reasoning. Moreover we observed women to employ strategies differently than men after extended instruction in the domain.

## Objectives & Theoretical Framework

A recent report from the National Research Council (2006) identifies spatial thinking as a critical component of scientific problem solving and reasoning and advocates for training spatial thinking in the science classroom. Such a call is consistent with the content of science instruction, which often requires students to reason about the three-dimensional relationships of objects and phenomena that are of interest to scientists. For example, chemistry students must learn about the three-dimensional structure of molecules, physics students must learn about the trajectory of projectiles and geology students must learn how geological structures transform over time. Given the prevalence of spatial thinking across the sciences, several researchers have suggested that student aptitude for spatial thinking, as measured by spatial ability psychometrics, predicts their success in science classrooms (Pallrand & Seeber, 1984; Wu & Shah, 2004) and careers (Shea, Lubinski, & Benbow, 2001). Indeed, a host of studies have shown positive correlations between visuo-spatial ability and achievement in several science domains (Carter, LaRussa, & Bodner, 1987; Hegarty & Sims, 1994; Keehner, Lippa, Montello, Tendick, & Hegarty, 2006). Consequently, these findings have led to claims that sex differences in spatial ability are responsible for sex differences in science achievement (cf. Fogg, 2005).

Despite the importance of visuo-spatial ability, questions remain about the cognitive components of spatial thinking. Typically, spatial thinking in science has referred to imagistic reasoning that includes mental imagery, mental rotation, spatial perspective taking and spatial visualization (Bodner & Guay, 1997). However, practicing scientists and novice students alike successfully solve spatial tasks through the use of external diagrams, models, and computer simulations that may or may not recruit these cognitive processes (Stieff, 2007; Stieff & Raje, 2010; Trafton, Trickett, & Mintz, 2005). Also, a variety of domain-specific analytic algorithms and heuristics have been reported that lead to solutions with little to no use of spatial information given in a spatial problem (Schwartz & Black, 1996; Stieff, 2007). The availability and utility of these alternative strategies raises several questions about the components of spatial thinking and their role in scientific problem solving at all levels.

The present paper aims to identify the underlying cognitive components that comprise spatial thinking in science. We address this aim with four questions: What strategies do problem solvers use to solve tasks that involve spatial thinking? Does strategy choice predict success on a variety of spatial tasks? Do spatial ability and sex predict strategy choice? How does instruction affect strategy choice? We address each of these questions by examining student problem solving in the domain of organic chemistry. Historically, this domain has privileged the role of visuo-spatial ability due to the content of organic chemistry which includes the analysis of three-dimensional relationships within and between molecular structures (Mathewson, 1999; Wu, Krajcik, & Soloway, 2001; Wu & Shah, 2004); yet, little is known about what strategies students employ when considering these relationships. Previously, Stieff and Raje (2010) have shown that expert chemists engage in spatial thinking using a variety of domain-specific diagrammatic and analytic strategies as opposed to mental imagery; however, strategy use among chemistry students remains unknown. Here, we build on the work of Stieff and Raje, by examining college students' choice of problem solving strategies for solving spatial organic chemistry problems to determine the extent to which chemistry students employ multiple strategies and how strategy choice changes with increasing domain knowledge.

## Study 1

In Study 1, we designed a strategy choice questionnaire that first asked students to solve 10 canonical organic chemistry assessment tasks. On each task, students were asked to indicate how they solved the problem using a list of known strategies applicable to the task. Previously, Stieff and Raje (2010) documented experts' use of specific imagistic and non-imagistic strategies for solving organic chemistry problems; the findings of that study were used to populate the list in the present work. The goal of Study 1 was to identify patterns of strategy use among students and any associations between strategy choice, achievement and sex.

## Method

Thirty-nine college students (20 males, 19 females) who had completed 6 months of instruction in organic chemistry were asked to complete a chemistry strategy choice questionnaire. The strategy questionnaire consisted of 10 organic chemistry problems that asked participants (1) to identify spatial relationships between molecules or substituents within a molecule and (2) to consider spatial transformations of molecular diagrams. All chemistry problems were scored for correctness using a binary rubric (1 = correct, 0 = incorrect). Participants were also asked to report the strategy they used to solve each chemistry problem by selecting from a list of possible strategies applicable to each problem. Participants were allowed to choose more than one strategy and to write in their own strategy if they believed that none of the choices matched their strategy. Each list of strategies for individual problems was developed in an earlier protocol study conducted by Stieff and Raje (2010); each strategy was coded according to a priori categories of strategy type listed in Table 1. Briefly, categories included those strategies that relied more extensively on reasoning via mental imagery (Spatial-Imagistic), diagrams (Spatial-Diagrammatic), rules and heuristics that operated on spatial information (Spatial-Analytic) and rules and heuristics that operated on non-spatial information (Algorithmic). Participants could also indicate if they knew the answer to a problem (Recall) or if they randomly guessed (Guessing). We note that the three categories that include the *spatial* prefix involve the direct consideration of spatial information while the algorithmic category does not. In cases where participants wrote in their own strategies, two researchers independently coded the free responses according to the four categories in Table 1. Comparison of the two raters' codes indicated an inter-rater reliability score above 85%.

Table 1: Strategy Categories.

| Strategy Type | Example Fixed-Choice Strategy Responses |
|---|---|
| Spatial-Imagistic | I tend to imagine the molecule in 3D and rotate it "in my head".<br>I tend to imagine myself moving into the paper or around the molecule. |
| Spatial-Diagrammatic | I tend to first draw a basic skeletal structure and then make changes as I go.<br>I tend to redraw the molecule using a different chemical representation to help me think about it. |
| Spatial-Analytic | I tend to assign R/S labels to each molecule. |
| Algorithmic | I just know that in stable molecules particular groups must be in a specific relationship.<br>I tend to use a specific formula to calculate the number of stereoisomers. |

## Results & Discussion

Figure 1 summarizes the frequency of each strategy choice across the 10 tasks. Among the 418 strategies reported, participants selected Spatial-Analytic strategies most frequently (36%) followed by Spatial-Diagrammatic strategies (26%), Spatial-Imagistic strategies (22%) and finally Algorithmic strategies (16%). Figure 2 shows a detail of strategy frequencies by task. The distribution of strategies differed dramatically among the ten tasks, which suggests that students freely switched between the different types of strategies depending on each task. For example, the majority of reported strategies applied to Tasks 1, 5, 6, and 8 were Spatial-Analytic strategies, but Spatial-Imagistic strategies were reported more often on Tasks 9 and 10. The dataset was further analyzed to determine whether participants used primarily one strategy for each task or applied multiple strategies. In total, we were able to identify the strategy used by participants in 326 (83.5%) of the 390 cases of problem solving. The remaining 64 cases either lacked strategy choice information or were solved via Guessing or Recall. As Table 2 illustrates, 240 tasks (73.6%) were solved with only one type of strategy, 80 tasks (24.5%) were solved with two types of strategies, and 6 tasks (1.8%) were solved with three or more types of strategies. In cases where participants used only one strategy, Spatial-Analytic strategies were reported most frequently. Interestingly, in cases where participants selected two types of strategies, the majority of reported strategies involved the use of a Spatial-Diagrammatic strategy and one other type of strategy. Notably, we observed a negative correlation between the number of participants who successfully completed a problem and the number of participants who used two or more strategies, ($r(10) = -0.654$, $p = 0.040$), which suggests that students tend to apply multiple types of strategies as questions become more difficult.

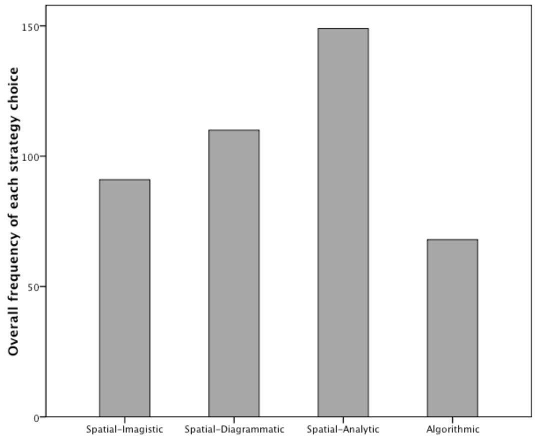



Figure 1. Overall frequency of strategies reported by category.

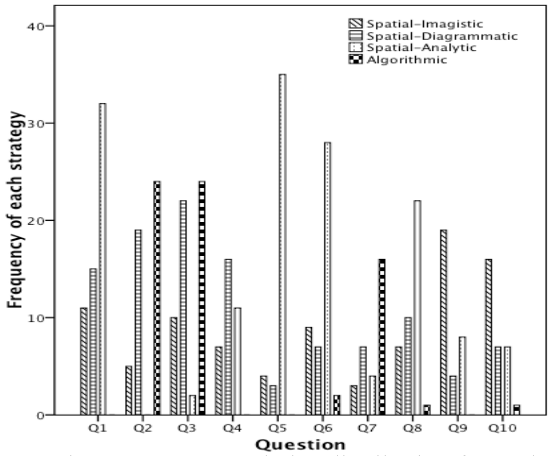



Figure 2. Strategy choice distribution for each task.

Table 2: Numbers and types of strategy used for each task.

| No. reported strategies used | Strategy type | Frequency | Total |
|---|---|---|---|
| 1 | Spatial-Imagistic | 49 | 240 (73.6%) |
| | Spatial-Diagrammatic | 47 | |
| | Spatial-Analytic | 107 | |
| | Algorithmic | 37 | |
| 2 | SI+SD | 16 | 80 (24.5%) |
| | SI+SA | 15 | |
| | SD+SA | 21 | |
| | SD+AL | 21 | |
| 3 or more | - | - | 6 (1.8%) |
| Total | | | 324 |

*Note.* Dashes indicate no further analysis was conducted. SI=spatial-imagistic, SD=spatial-diagrammatic, SA=spatial-analytic, AL=algorithmic. 64 tasks coded as recall/guessing/unknown are not included.

The relationship between correctness and type of strategy used was tested using a Pearson's $\chi^2$ test for 2 (use of each strategy) x 2 (correctness) contingency table. Using an alpha level of 0.05, no association between success and strategy use was found, indicating that strategy choice does not have an impact on whether a participant answer a task correctly. Sex differences in problem-solving success and strategy choice were tested using an independent two-sample $t$-test. The mean total correctness score of male participants ($M = 4.25$, $SD = 1.45$) was not found to differ from the mean total correctness score of female participants ($M = 4.15$, $SD = 1.12$), $t(37) = 0.22$, $p = 0.41$. Likewise, strategy choice did not differ significantly between female and male participants, as illustrated in Figure 3. Men and women displayed similar patterns of strategy choice: in order of reported strategy use, both groups employed Spatial-Analytic, Spatial-Diagrammatic, Spatial-Imagistic and Algorithmic strategies. In order to examine the relationship between sex and strategy choice, strategy scores of participants were calculated by counting the numbers of each strategy used across the ten survey items. $t$-tests to compare Spatial-Imagistic, Spatial-Diagrammatic, Spatial-Analytic and Algorithmic strategy scores between male and female were not found to be statistically significant at an alpha level of 0.05.

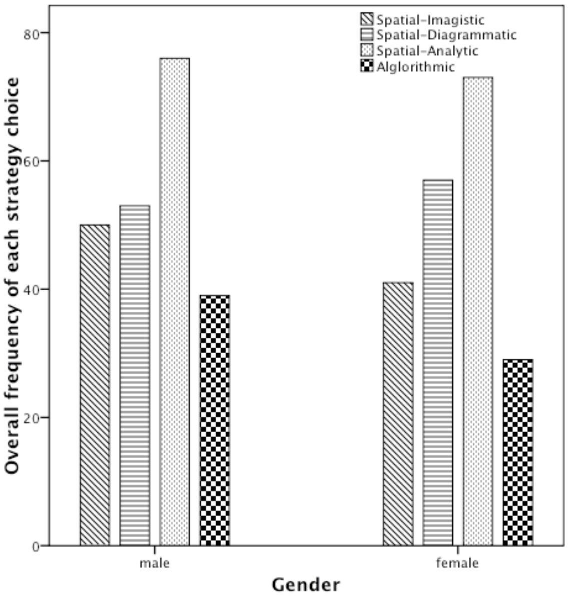



Figure 3. Overall frequency of strategy choice by males and females.

## Study 2

In Study 2, we adapted the strategy choice questionnaire for group administration via a remote personal response system (i.e., "clickers") in an organic chemistry classroom during instruction. Although Study 1 established that students primarily made use of spatial-analytic strategies for solving organic chemistry tasks, the participants in that studied had completed several months of instruction in the domain. Thus, Study 1 yielded no information about how student strategy choice changes with instruction. Therefore, we conducted Study 2 to determine whether students employed spatial-analytic strategies in the context of an organic chemistry course and whether students employed the same strategies uniformly over the course of instruction.

## Method

103 undergraduate students enrolled (sex was reported for 90 students: 33 males and 57 females) in a 6-week intensive organic chemistry course were assigned unique personal response devices to respond to adapted strategy choice questions administered during the course. Over the duration of the course, students were asked 10 unique organic chemistry questions and related strategy choices. Questions were administered approximately once each week of instruction. During the final meeting of the course, students were asked 8 organic chemistry questions and related strategy choices that included 6 of the 10 questions administered during earlier sessions of the course. All questions were presented on large LCD televisions at the front of the classroom and students answered questions by clicking a multiple-choice answer on their assigned device. The scoring rubric and strategy categories from Study 1 were used to analyze student responses. Notably, the adapted questions in Study 2 did not contain algorithmic strategies as the course instructor deemed that the strategy survey questions that included algorithms were beyond the scope of her course. In addition, unlike the strategy survey questionnaire, students were not able to choose more than one strategy per problem because the classroom clicker system could not capture multiple answers per student for a given question. Students were able to report their own strategies after each class if they employed a strategy not presented in the provided options.

Among the 103 students, 91 students volunteered to complete a spatial ability battery that included the Vandenberg Mental Rotation Test (Vandenberg & Kuse, 1978) and Guay's Visualization of Views (McDaniel & Guay, 1976). Descriptive statistics of strategy choice were generated for each task and strategy use on both administrations of the 6 questions was compared. Unlike Study 1, group administration of the questions permitted students to interact and discuss their responses prior to inputting an answer on their clicker devices and the course instructor assigned these questions for course credit; therefore, the independence of student answers to chemistry problems could not be guaranteed and reports of student achievement were not considered valid for analysis. In contrast, because students did not receive credit for strategy responses and the instructor emphasized that there was no correct answer to these questions, we considered student responses to these questions valid for analysis.

## Results & Discussion

The distribution of strategy choices at each administration time point in the classroom is presented in Table 3. As indicated, the students reported that they employed Spatial-Imagistic strategies more than any other strategies both during and after instruction. Excluding recall, guess, and unreported strategies, Spatial-Imagistic strategies were most frequently reported by students (947 times, 64.95%), followed by Spatial-Diagrammatic strategies (397 times, 27.23%) and Spatial-Analytic strategies (114 times, 7.82%). Although Spatial-Imagistic strategies dominated both during and after organic chemistry instruction, comparison between the two occasions suggests that fewer Spatial-Imagistic strategies were employed after instruction while Spatial-Diagrammatic and Spatial-Analytic strategies were reported more frequently.

Reports of strategy use on the six questions appearing both during and after instruction were examined further to clarify changes in strategy use after instruction. As indicated in Table 4, after instruction the average number of Spatial-Imagistic strategies across all tasks reported decreased ($t(102) = -3.98$, $p < .001$), and the average number of Alternative strategies increased ($t(102) = 4.95$, $p < .001$). Figure 4 illustrates the frequency of reported strategies for each of the six questions at each presentation. Examination of these items indicates that students do indeed employ Spatial-Imagistic strategies less frequently after instruction. Interestingly, distributions of strategy choice after instruction varied across the six question items. For questions 1 and 6, reports of using Spatial-Analytic strategies rose dramatically, while reports of using Spatial-Diagrammatic strategies rose relatively higher on questions 2 and 4. In contrast, no noticeable difference in the relative use of each strategy type was seen on questions 3 and 5. Examination of these six items revealed that students not only adopted strategies alternative to Spatial-Imagistic Strategies after instruction, but the choice of strategy after instruction was related to the task itself.

Table 3: Frequency of strategy use.

| | No. strategy choice | | |
|---|---|---|---|
| Types of strategy | During instruction [a] | After instruction [a] | Total |
| Spatial-Imagistic | 596 (73.22%) | 351 (54.50%) | 947 (64.95%) |
| Spatial-Diagrammatic | 172 (21.13%) | 225 (34.94%) | 397 (27.23%) |
| Spatial-Analytic | 46 (5.65%) | 68 (10.56%) | 114 (7.82%) |
| Total | 814 (100 %) | 644 (100%) | 1458 (100%) |

[a] 10 question items were administered during instruction and 8 items were administered after instruction.

Table 4: Mean number of Spatial-Imagistic and Alternative strategies reported during and after instruction.

| | During the instruction | | After the instruction | |
|---|---|---|---|---|
| Types of strategy | *M* | *SD* | *M* | *SD* |
| Spatial-Imagistic strategies | 3.56 | 1.48 | 2.63 | 1.91** |
| Alternative strategies | 0.99 | 1.09 | 1.83 | 1.79** |

*Note*. Scores for each category range from 0-6 excluding recall and guessing strategies.
** $p < 0.001$

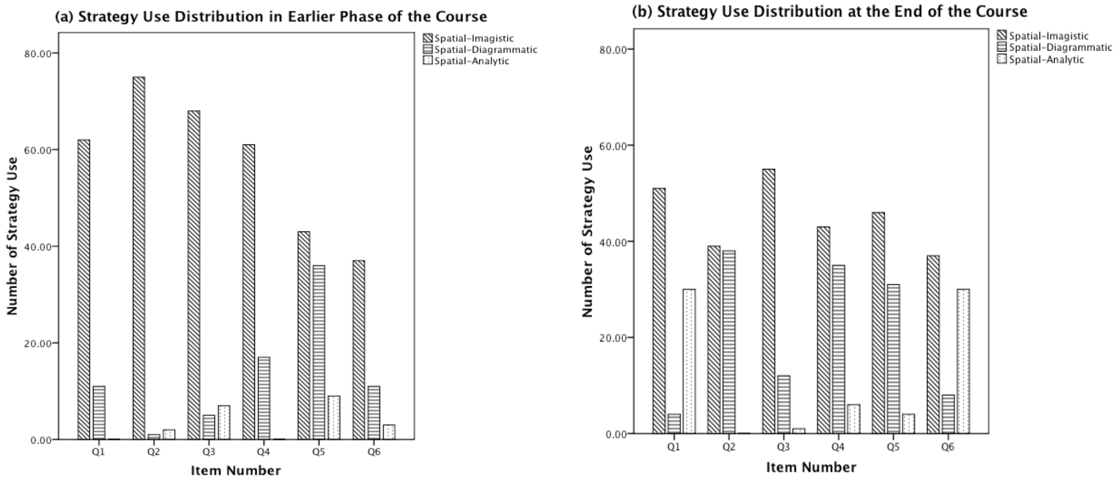



Figure 4. Frequency of strategy use reported by students (a) during and (b) after instruction.

Associations between spatial ability and strategy choices were analyzed via ANOVA. The 91 students who completed the spatial ability psychometrics were categorized into three groups based on their performance on the Mental Rotation Test (MRT) and Visualization of Views Test (VoV): High ($N$=31, $M$=51.94, $SD$=13.02 for MRT and $M$=17.74, $SD$=4.56 for VoV), Medium ($N$=30, $M$=34.20, $SD$=10.39 for MRT and $M$=8.71, $SD$=4.44 for VoV), and Low ($N$=30, $M$=15.40, $SD$=12.08 for MRT and $M$=4.21, $SD$=3.58 for VoV). Table 5 illustrates the results from the ANOVA. On the first presentation of each strategy question, the use of Alternative strategies did not vary with spatial ability ($F(2, 88)$=0.96, ns) at an alpha level of 0.05. After instruction, however, we observed a trend in the data that indicated students in the lower ability group employed Alternative strategies more frequently than higher spatial ability students ($F(2, 88)$=3.10, $p$=0.05). Associations between each student's strategy choice and spatial ability were analyzed via a Multivariate Analysis of Variance (MANOVA) test of Alternative strategy scores with within-subjects effect of administration time (i.e., during and after the instruction) and between-subjects effect of spatial ability group. The analysis failed to show a significant interaction between student strategy choice after instruction and spatial ability, *Wilk's* $\lambda$= 0.968, $F$ (2, 88)=1.43, ns. Thus, spatial ability was not found to predict the use of any particular strategy after instruction.

Table 5: Comparison of Alternative scores during and after the instruction in three spatial ability groups

| | High | | Medium | | Low | | |
|---|---|---|---|---|---|---|---|
| Occasions of the task | *M* | *SD* | *M* | *SD* | *M* | *SD* | $F$ (2,88) |
| During the instruction | 0.90 | 1.07 | 1.30 | 1.08 | 1.06 | 1.20 | 0.96 |
| After the instruction | 1.35 | 1.56 | 2.43 | 1.94 | 2.17 | 1.80 | 3.10 |

Finally, relationships among sex, spatial ability and strategy choice were investigated. Using an alpha level of 0.05, males were found to outperform females on the Mental Rotation Test ($M$ = 45.63, $SD$ = 16.07 for male, $M$ = 27.62, $SD$ = 17.63 for female, $t(83)$ < 0.001) and on the Visualization of Views ($M$ = 13.42, $SD$ = 8.27 for male, $M$ = 8.66, $SD$ = 5.94 for female, $t(83)$ = 0.003). During instruction, males and females did not differ in use of Spatial-Imagistic strategies ($M$ = 3.81, $SD$ = 1.36 for male, $M$ = 3.32, $SD$ = 1.40 for female, ns) or the use of Alternative strategies ($M$ = 1.00, $SD$ = 0.87 for male, $M$ = 1.07, $SD$ = 1.10 for female, ns). After instruction, however, females were observed to use Alternative strategies more frequently than males ($M$ = 1.24, $SD$ = 1.56 for male, $M$ = 2.47, $SD$ = 1.79 for female, $t(88)$ = 0.002); however, the difference between male and female use of Spatial-Imagistic strategies was marginal ($M$ = 3.21, $SD$ = 1.92 for male, $M$ = 2.53, $SD$ = 1.83 for female, $t(88)$ = 0.096). Repeated measures analysis of Alternative strategy scores involving within-subjects effect of administration time (i.e., during and after instruction) and between-subjects effect of sex resulted in significant interaction between sex and occasion of the tasks (MANOVA, *Wilk's* $\lambda$ = 0.892, $F(1, 88)$ = 21.31, $p$ = 0.002).

## Conclusions & Implications

The above results offer some tentative answers to the questions we posed initially. First, the findings clearly illustrate that students employ a variety of strategies to solve tasks that involve spatial thinking. In Study 1, we observed students to rely more consistently on Spatial-Analytic and Spatial-Diagrammatic strategies as opposed to Spatial-Imagistic strategies, as typically believed. Likewise, in Study 2, we observed students to employ Spatial-Imagistic strategies preferentially during instruction, yet adopt more alternative strategies by the end of the course. Moreover, we also observed students to fluidly switch between different types of strategies between tasks. The findings of the present work suggest that students choose task-dependent strategies in a manner similar to expert chemists and apply multiple strategies on problems of increased difficulty. These results indicate that students are aware of the availability of diverse strategies and are willing to employ alternative strategies. In other words, students are not limited to reasoning about spatial information in molecular structures via imagistic reasoning, but can reason about spatial information with a variety of strategies.

The results also indicate that strategy choice does not predict success on spatial tasks in chemistry. The findings in Study 1 suggest that students reach equivalent levels of achievement regardless of whether they employ strategies that involve reasoning via mental imagery or alternative strategies. Equally important, we did not observe significant differences in achievement between men and women on chemistry tasks. Despite these findings, we did observe that multiple strategies were applied on tasks that the majority of students failed to solve. This finding is consistent with the literature on flexible strategy choice that reports individuals employ multiple strategies on tasks of increased difficulty (cf. Siegler, 1996). The use of multiple strategies, however, did not lead to increased success on such tasks. Thus, it did not appear that the application of one or more strategy types (e.g., Spatial-Imagistic, Spatial-Analytic, Spatial-Diagrammatic, Algorithmic) predicts achievement. That is, each strategy is equally likely to result in success or failure on a given task.

Study 2 permitted us to examine the relationship between strategy choice and instruction in the context of an organic chemistry classroom. The results of that study clearly illustrate that instruction has a direct effect on strategy choice. In the beginning of the course, we observed students rely primarily on Spatial-Imagistic strategies to solve spatial tasks; by the end of the course, we observed a sharp increase in the use of strategies alternative to Spatial-Imagistic strategies. Interestingly, the participants in Study 2 reported greater use of Spatial-Imagistic strategies at the end of instruction while the participants in Study 1 reported greater use of Spatial-Analytic strategies. We believe the reason for this discrepancy is due to two important differences between the participants in each study. First, the students received instruction over different time periods. The students in Study 1 completed ~20 weeks of instruction during course of an academic year; however, the students in Study 2 learned less material in a 6 week summer course. It is possible that the longer duration of study in Study 1 resulted in better apprehension of and preference for alternative strategies. Similarly, the instructors for each course reported notable differences in their own emphasis on strategy use. The instructor in Study 1 reported she was 'bad at visualization' and emphasized diagrammatic and algorithmic heuristics, but the instructor in Study 2 reported she attempted to teach as many strategies as possible for the benefit of the students. Thus, instructional differences may have resulted in the observed differences in strategy preference. Nevertheless, although students in Study 2 reported using Spatial-Imagistic strategies as their primary strategy, the increased use of domain-specific alternative strategies suggests that as expertise develops, students may rely less on imagistic reasoning and more on heuristics to solve spatial tasks.

Study 2 also permitted us to examine the relationship between spatial ability, sex and strategy choice in the classroom. Although the results of that study do not indicate a direct relationship between spatial ability, sex and strategy choice, they do suggest a potential interaction may exist. First, our findings clearly show that over the course of instruction women reported a significant increase in the use of alternative strategies compared to men. Second, our findings tentatively suggest that low spatial students may preferentially switch from Spatial-Imagistic strategies to alternative strategies after instruction; high spatial students do appear to rely on Spatial-Imagistic strategies throughout instruction. Thus, the data suggests that low-spatial females preferentially switch to alternative strategies. Two major limitations of the Study limit the validity of these findings. First, our analysis relies solely on students' strategy reports on 6 questions. The results of Study 1 indicate that several strategies are task-specific and our reliance on so few tasks casts doubt on the interpretation of these findings. Second, students were asked to respond to the clicker questions in Study 2 under classroom time constraints and they were also permitted to collaborate on their responses. Thus, there was an increased risk in Study 2 of failing to detect changes in strategy choice and individual differences in spatial ability. Nevertheless, we believe the trends in the data suggest a potential interaction between spatial ability, sex and strategy choice does exist and warrants further investigation.

Taken together, the results of the present studies indicate that spatial thinking in advanced scientific problem solving, specifically organic chemistry, involves a range of strategies that vary significantly in the extent to which they rely on imagistic reasoning. Of particulate note, our findings suggest that students approach the study of organic chemistry using mental rotation and other spatial-imagistic strategies to reason about molecular structures, but quickly adopt a variety of algorithms and heuristics after instruction. This behavior leads us to question the utility of instructional methods that emphasize the exclusive focus on training students to use imagistic strategies (e.g., by improving students' visuo-spatial ability, Ferk, Vrtacnik, Blejec, & Gril, 2003). Rather, we suggest instead that students may benefit most from instruction that teaches the applicability of multiple strategies, as in Study 2. Moreover, the present study did not identify significant correlations between sex and chemistry problem solving success. This result contradicts previous claims that men outperform women in science due to their aptitude for spatial reasoning (Fogg, 2005). Rather, our findings suggest that female students apply the same strategies as male students with equal levels of success in chemistry and that they are likely to switch to alternative strategies when necessary in a course.

## References

Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Educator, 2*(4), 1-18.

Carter, C. S., LaRussa, M. A., & Bodner, G. M. (1987). A study of two measures of spatial ability as predictors of success in different levels of general chemistry. *Journal of Research in Science Teaching, 24*(7), 645-657.

Ferk, V., Vrtacnik, M., Blejec, A., & Gril, A. (2003). Students' understanding of molecular structure representations. *International Journal of Science Education, 25*(10), 1227-1245.

Fogg, P. (2005, January 28). Harvard's president wonders aloud about women in science and math. *The Chronicle of Higher Education*, p. A12.

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22*(4), 411-430.

Keehner, M., Lippa, Y., Montello, D. R., Tendick, F., & Hegarty, M. (2006). Learning a spatial skill for surgery: How the contributions of abilities change with practice. *Applied Cognitive Psychology, 20*, 487-503.

Mathewson, J. H. (1999). Visual-spatial thinking: An aspect of science overlooked by educators. *Science Education, 83*, 33-54.

National Research Council (2006). *Learning to think spatially*. Washington, D.C.: The National Academy Press.

McDaniel, E. D., & Guay, R. B. (1976). *Spatial ability, mathematics achievement, and the sexes*. Paper presented at the Annual Meeting of American Educational Research Association, San Francisco, CA.

Pallrand, G. J., & Seeber, F. (1984). Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching, 21*(5), 501-516.

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science, 20*, 457-497.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93*(3), 604-614.

Siegler, R. S. (1996). *Emerging minds*. New York: Oxford University Press.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction, 17*, 219-234.

Stieff, M. & Raje, S. (2010). Expertise and imagistic reasoning in chemistry. *Spatial Cognition & Computation*, *10*(1), 53-81.

Trafton, J. G., Trickett, S. B., & Mintz, F. E. (2005). Connecting internal and external representations: Spatial transformations of scientific visualizations. *Foundations of Science, 10*, 89-106.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotation: Group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-604.

Wu, H.-K., Krajcik, J. S., & Soloway, E. (2001). Promoting conceptual understanding of chemical representations: Students' use of a visualization tool in the classroom. *Journal of Research in Science Teaching, 38*(7), 821-842.

Wu, H.-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education, 88*(3), 465-492.

## Acknowledgments

This work was supported, in part, by a grant from the National Science Foundation (DRL-0723313). Any opinions, findings or conclusions expressed in this paper are those of the author and do not necessarily represent the views of this agency.

# What Are They Talking About? Findings from an Analysis ofthe Discourse in Peer-Led Team Learning In General Chemistry

Patrick Brown, Keith R. Sawyer, Regina Frey, Sarah Luesse, Daniel Gealy, Washington University in St. Louis, Campus Box 1022, One Brookings Drive, Eads Hall Rm. 105
Email: Pbrown23@wustl.edu, ksawyer@wustl.edu, gfrey@wuchem.wustl.edu

**Abstract:** Peer-Led Team Learning (PLTL) is a structured method for helping students engage actively in collaborative conversations. The method originated in undergraduate chemistry courses, but is now used in math and in other science classes as well. Previous studies have shown that PLTL results in improved student learning in undergraduate chemistry. However, researchers have not studied the group mechanisms and the discourse processes that lead to this improved outcome. This study is one of three inter-related studies that are the first to explore those mechanisms and processes. In this study, we observed videotapes of PLTL sessions and analyzed the discourse of peer leaders and of students. We found that the structure and nature of the problems influences student discourse.

## Introduction

There is now a consensus in science education research that the most effective learning environments are those in which students engage in productive, collaborative discourse to build knowledge (e.g., American Association for the Advancement of Science [AAAS], 1989; National Research Council [NRC], 1996). Knowledge building occurs when students engage in collaborative conversations intended to advance both individual understanding and the collective knowledge of the group in pursuit of a common goal (Bereiter, 2002; Engle & Conant, 2002; Rogoff, Matusov, & White, 1996). In response to this research, numerous institutions of higher education have introduced a form of collaborative learning into lower level science courses that is called peer-led team learning (PLTL) (Gafney & Varma-Nelson, 2008; Gosser et al., 2001; Gosser & Roth, 1998; Hockings, DeAngelis, & Frey, 2008; Sarquis et al., 2001). Although there are still large lectures each week, PLTL supplements the lecture with formalized study groups that are facilitated by a peer leader and provide opportunities for active and collaborative learning. A PLTL study group contains 6-8 students facilitated by a peer leader, a student who has previously received a high grade in the class, and who works under the close supervision of the instructor of the class. In each PLTL session, students work together to solve problems designed by the instructor of the class. Neither the peer leader nor the students are given the solutions to the problems, because the goal of the session is not to get the correct answer; instead, it is to provide opportunities for engaging in problem solving while discussing the concepts in the problem. The goals of PLTL are to: (i) teach undergraduates how to effectively study in a group; (ii) improve students' problem-solving skills; (iii) provide facilitated help for students; and (iv) provide an active- learning environment for students to engage in scientific discourse.

Although previous studies have shown that PLTL results in improved learning in undergraduate chemistry (Gafney & Varma-Nelson, 2008; Hockings, DeAngelis, & Frey, 2008; Tien, Roth, & Kampmeier, 2002), few researchers have studied the group mechanisms and discourse processes that lead to this improved outcome. In two inter-related studies, we found that peer leaders used two distinct interactional styles, which we call *instructional* and *facilitative*, that influence students' dialogue, participation, and knowledge building. In the first study, we found that peer leaders who use a high percentage of facilitative discourse was related to increased chains of student-to-student interactions and more equal student participation (Brown, Sawyer, & Frey, 2009b). Conversely, peer leaders who used equal amounts of facilitative and instructional discourse had shorter chains of interactions and unequal participation. In the second study, we observed that when a peer leader used primarily facilitative discourse and provided abundant managerial support, students displayed extended discussions that went beyond applying equations in a rote manner and began to develop an understanding of the concepts (Brown, Sawyer, & Frey, 2009a). This study is the third in a series of studies that examines the discourse practices used in PLTL. After observing the transcripts of the first two studies, we believed that the specific nature and framing of the problem presented to the students had an impact on their discourse, and we wanted to better understand the specific links between the posed problem and the resulting discourse. Thus, the research question we posed for this study is: How does the nature and structure of the problem influence how students talk about the underlying concepts?

# Research Design

## Participants

In fall 2006, during the 14-week semester, three PLTL sessions (# 4, #7, and #9) of each of 15 returning peer leaders were videotaped. For the project, all verbal interactions in 6 PLTL groups for two parts of session 7 were analyzed from these recordings. We used a purposeful sampling approach to identify cases and parts of problems that were "information-rich" (Beals & Tabors, 1995; Patton, 2002). The 6 groups were chosen based on observation of all the video data. Based on our experiences working with peer leaders, we purposefully selected: (1) 6 PLTL groups that represented a range of peer leader styles (instructional or facilitative) (see Brown, Sawyer, & Frey, 2009b); and (2) questions that represented typical PLTL activities (reviewing content, problem solving, and concept discussion).

## Data Analysis

We implemented multiple approaches to analyze the data. We identified, labeled, and time-stamped the amount of peer-leader talk, student talk, individual tasks, and off-task behaviors that occurred during the selected PLTL activities based on relevant, transcribed portions of the video data. Each transcription was divided into segments of talk, in which a segment of talk represented an individual's contribution to the discussion. An individual's talk could consist of multiple utterances depending on how many ideas were included in one segment of talk. Each utterance was assigned a code (Brown, Sawyer, & Frey, 2009b). The coding scheme involved two levels of coding: categories and codes. The categories we developed were consistent with the constant comparative method of qualitative data analysis (Glaser & Strauss, 1967). Categories are general types of speech observed and include 4 broad areas: (1) explanation, (2) content question, (3) facilitation, and (4) problem solving. Codes refer to specific types of identifiable speech within these 4 categories. For example, closed question and open question are both codes within the category "content question."

To calculate the reliability of the coding manual, two trained coders were used; one has a background in science education and qualitative research and the other was a co-instructor of the general chemistry series. The second coder was blind to the motivation and hypotheses of the study. When disagreements arose, the two coders discussed differences and either a rule for coding was decided upon or a revision was made to the coding manual. Cohen's Kappa is an inter-rater reliability measure for qualitative studies (Bakeman & Gottman, 1986; Lunn, 1998). Bakeman and Gottman (1986) characterized a Cohen's Kappa of greater than 0.75 as excellent. The Cohen's Kappa for the final stage of coding was 0.90; thus, meeting the criteria for excellent inter-rater reliability.

# Results

Our data at this stage of the project is preliminary. Thus, we chose relevant portions of the data from our qualitative analysis to highlight findings from our full data set. We focus on a specific trend in our data that we observed across our 6 PLTL groups.

To study whether the nature and structure of the problem influences knowledge-building discourse we examined a portion of the seventh PLTL session. During the session, students reviewed and solved problems associated with understanding the topic "periodic trends" — characteristics of elements as one moves across a period (row) and down a group (column) of the periodic table. This analysis focuses on the Review and Problem 1 portion of the session. During the review, students discussed the content covered in lecture and recitations. During Problem 1, students arrange each set of atoms (e.g., (a) Rb, Cs, Li; (b) B, Li, F; (c) Cl, F, Br; (d) Rb, Be, K) in order of increasing atomic radius.

## Review

The transcript that follows is an example of how Gillian's students began to talk about the periodic trends during the review.

[1] F1: So I have a bunch of periodic trends (inaudible). There's Z* (effective nuclear charge), one for electronegativity, and one for Atomic radius. Do you wanna draw all those?
M7: And electron affinity.
F1: Draw the, Draw the boxes [[2]*Student tells peer to draw a "box," in the shape of a rectangle, on the*
F4: *front board*]
F1: Why don't you just draw the boxes?
F4: I like the boxes. (laughs). Just write like, Z*, and then arrow, increasing (pointing). And then arrow down, decreasing (pointing). [*Student tells her peer who is at the front board to draw a box with a one-sided arrow pointing from left to right and another arrow pointing from the bottom to the top of the box*]

[1]Line number
[2]Relvant nonverbal interactions

In this excerpt, students drew a “box” to represent the periodic table and used arrows to indicate whether atoms increase or decrease across a period (row) and down a group (column) according to the different diagrams for the period trends that they covered in lecture (e.g., electronegativity, atomic radius, electron affinity, and Z*) (LN: 1-6). For example, students began talking about the periodic trends by focusing on the procedures for describing how Z* increases and decrease across and down the periodic table (LN: 7). When identifying important content to review from the lectures and recitations, and without prompting from the peer leaders, students used the diagrams they had learned in class as an algorithm (a set of finite rules) for solving problems (LN: 7). Without the peer leaders guidance, we found that students used the diagrams they had learned in class almost exclusively when reviewing the periodic trends.

## Problem 1

In problem 1, students arranged atoms in order of increasing atomic radius. The problem-solving discourse was characterized by students using the diagrams they discussed during the review to solve atomic radius problems. The transcript that follows illustrates how Rachael’s students focused on applying the diagram that was discussed during the review to arrange the atoms B, Li, and F in order of increasing atomic radius. Rachael’s students’ discourse was typical of how students talked about the atomic radius problem (Problem 1).

254 M1: Umm, Bi, Li, and F (reading the problem). So we’re going left to right, and the diagram says is
255 decreases (referring to the diagram on the board developed during the review), so, if I want it
256 from increasing, it’s Li, B, F. [*Student looks at the diagram on the board showing a one-sided*
257 *arrow pointing from right to left to answer the question*]
258 F3: No
259 M2: No, it goes the other way (inaudible).
260 PL: Increasing should start with the smallest, and then…
261 M1 Oh. So, F, B, Li.
262 PL: Don’t let it trip you up. Does everyone agree?
263 F3: Yeah.

In the excerpt above, M1 used the diagram on the board that they discussed during the review to solve the problem (LN: 254). In this example, M1 accidentally listed the elements from largest to smallest atomic radii (the directions in the problem asked students to list the elements from smallest to largest atomic radii). In line 259, the peer leader calls to attention that M1 has placed the elements in the reverse order. In line 260, M1 reacts by listing the elements in the correct order. In this excerpt, students discourse was entirely focused on the diagram discussed during the review and they did not discuss the underlying concepts associated with the trend.

Gillian’s students’ discourse was also focused on applying the diagram they had discussed during the review on problem 1. However, Gillian engaged students in the concepts by using probing questions (LN: 337).

337 PL: …So what's the, what's the idea with radius? What are you doing, to kind of put those(referring to
338 atoms) in order?
339 F1: We just compared it to the, the, drawing up there (referring to the front board), the table.
340 F5: The number of, like, protons, um, of each element and then the number of, I don’t know, I guess
341 electrons?...
342 F1: Well, one, we didn't have any ions, so that wasn't really important (laughs).
343 F5: Right. So if there was like, more protons, like, going down the periodic table, like down a column,
344 or whatever, um, they (referring to atomic radii) would be increasing because they would just be a
345 bigger
346 F1: Well, there's more elect...(referring to electrons)
347 F5: It would be more electrons...If we're going down, it's, it would be more electrons.
…[3] … …
420 PL: So you guys down there, what's the, what's kinda like the trend?
421 M7: Whichever one has the lower number of protons has a tighter pull on the electrons, so the radius
will be smaller (across a group)
… … …
425 PL: … (Does everyone) agree with that?
426 F1: And as you're adding elect..., it's like the same if you're adding electrons (down a period), then an
427 increased radius.

[3]Transcripts were omitted because student talk was off task.

Gillian's students' responses indicate that they focused on applying the diagram they had discussed for atomic radius during the review to the problem (see line 339). However, Gillian used probing questions to promote discussions, student interactions, and scientific explanations of the content (LN: 337 and 420). Students acknowledged, built upon, and elaborated on each other's ideas. Additionally, Gillian's students' explanations went beyond using the diagrams they learned in class to solve the problems and they began to address some of the underlying concepts (LN: 340-346, 421, and 426).

In summary, both the nature and structure of the problem influenced student discourse. We found that during the review students focused on describing the periodic trends according to diagrams that they learned about in class. Without prompting from the peer leader, students almost exclusively talked about the periodic trends according to a finite set of rules that they learned about in class. Students' discussion during the review had an influence on how they talked about the periodic trends problems. Students used the diagrams that they discussed during the review to solve the periodic trends problems. Thus, the nature and the structure of the PLTL problems influenced students discourse.

## Conclusion

There is a great deal of research evidence that students who participate in PLTL perform better than students who learn individually and alone. Our work begins to examine exactly how the discourse in PLTL groups contributes to improved chemistry understanding. From our analysis, it appears that not all peer-group experiences are equivalent in promoting chemistry understanding.

Based on previous research we found that peer leaders who used equal amounts of facilitative and instructional discourse had shorter chains of interactions and unequal participation. Conversely, the use of a high percentage of facilitative discourse was related to increased chains of student-to-student interactions and more equal student participation (Brown, Sawyer, & Frey, 2009b). Second, we observed that when a peer leader used primarily facilitative discourse and provided abundant managerial support, students displayed extended discussions that went beyond applying equations in a rote manner and began to develop an understanding of the concepts. Conversely, when a peer leader used more equal combinations of discourse coded as instructive and facilitative, students spent considerable time working individually and their discourse focused on the algebraic steps necessary to solve the problem (Brown, Sawyer, & Frey, 2009a). In this study, we observed that students' discourse was related to the nature and structure of the problem. Students mostly talked about the periodic trends using the diagrams they learned about in class. The discussions that occurred during the review influenced how the students solved periodic trends problems. Students focused on using the diagrams they discussed during the review to solve problems and rarely engaged with the underlying concepts associated with understanding the periodic trends. Thus, for some problems, students would benefit from explicit conceptual questions in order to engage in discussions of the underlying concepts.

The goal of PLTL is to engage students in building chemistry knowledge through a more open forum of collaborative discourse with greater student involvement. To provide quality PLTL sessions, Chemistry instructors must understand the discourse practices of peer leaders and students. Many education researchers have stressed that an effective student culture of collaboration requires close attention and scaffolding, and once a collaborative group culture has emerged, classroom conversations can lead to significant individual cognitive advancement and the development of deep conceptual understanding (Greeno, 2006; Sawyer, 2006; Scardamalia & Bereiter, 2006). This preliminary study suggests that the nature and structure of the problem influence student discourse. If problems include explicit conceptual questions then collaborative group cultures could emerge that encourages students to engage in deep knowledge building.

## Implications for Future Research

Examining student discourse during a content review and problem solving activity has implications for the redesign of PLTL problems and future research. We believed that students would discuss the underlying concepts associated with the review and the problem based on their experiences in lecture and recitation sessions. Based on the early stages of our third study, implications arise for the redesign of PLTL problem sets. Based on the findings of the third study we are redesigning PLTL problem-sets to begin with guiding questions that encourage students to discuss key concepts and experiments in addition to equations and variables. Altering existing problem sets to provide explicit questions that have students discuss phenomena before problem solving may engage students in higher-order thinking and alter students' interactions with each other and the peer leader. Future research is needed that investigates whether revising the way PLTL problems are written better fosters the type of conversations that lead to deep conceptual understanding. Restructuring the problems could favorably affect student's chemistry understanding, critical thinking, and knowledge building from collaborative discourse.

## References

Bakeman, R., & Gottman, J. (1986). Observing interaction: An introduction to sequential analysis. New York: Cambridge University Press.

Beals, D. E., & Tabors, P. O. (1995). Arboretum, bureaucratic, and carbohydrates: Preschoolers' exposure to rare vocabulary at home. *First Language, 15*, 57-76.

Bereiter, C. (2002). *Education and mind in the knowledge*. Mahwah, NJ: Erlbaum.

Brown, P., Sawyer, K. R., Frey, R. (2009a, Oct). *Peer-led team learning in general chemistry: Investigating the discourse of peer leaders and students.* Paper presented at the annual meeting of the Mid-western educational Research Association Conference, St. Louis, MO.

Brown, P., Sawyer, K. R., Frey, R. (2009b, April). *Investigating peer-leader discourse in peer-led team learning in general chemistry.* Poster presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition & Instruction, 20*(4), 399.

Gafney, L., & Varma-Nelson, P. (2008). *Peer-led team learning: Evaluation, dissemination, and institutionalization of a college level initiative.* Dordrecht, the Netherlands: Springer.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Gosser, D. K., Cracolice, M., Kampmeier, J., Roth, V., Strozak, V., & Varma-Nelson, P. (2001). *Peer-Led Team Learning: A guidebook*. Upper Saddle River, NJ: Prentice Hall.

Gosser, D. K., & Roth, V. (1998). The workshop chemistry project: Peer-led team learning. *Journal of Chemical Education, 75*(2), 185-187.

Greeno, J. G. (2006). Learning in Activity. In R. K. Sawyer (Ed.), *The cambridge handbook of the learning sciences* (pp. 79-96). New York: Cambridge University Press.

Hockings, S. C., DeAngelis, K. J., & Frey, R. F. (2008). Peer-led team learning in general chemistry: Implementation and evaluation. *Journal of Chemical Education, 85*(7), 990-996.

Lunn, M. (1998). Applying simple k-sample tests to conditional probabilities for competition risks in a clinical trial. *Biometrics, 54*, 1662-1672.

Patton, M. Q. (2002). *Qualitative research and evaluation* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Rogoff, B., Matusov, E., & White, C. (1996). Models of teaching and learning: participating in a community of learners. In D. R. Olson & N. Torrance (Eds.), *Handbook of education and human development* (pp. 338-414). Malden, MA: Blackwell Publishing.

Sarquis, J. L., Dixon, L. J., Gosser, D. K., Kampmeier, J. A., Roth, V., Strozak, V. S., et al. (2001). The workshop project: Peer-led team learning in chemistry. In J. E. Miller, J. E. Groccia & M. S. Miller (Eds.), *Student-Assisted Teaching* (pp. 150-156). Boston: Anker.

Sawyer, R. K. (2006). Analyzing collaborative discourse. In K. R. Sawyer (Ed.), *The Cambridge Handbook of the Learning Science* (pp. 187-204). New York: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97-118). New York: Cambridge University Press.

## Acknowledgements

The research in this paper has been funded by NSF CCLI grant award #0633202. The authors thank all of the peer leaders for their efforts in making the PLTL program successful at Washington University in St. Louis

# A Web-based Reading Environment Designed to Fundamentally Extend Readers' Interaction with Informational Texts

**Abstract**: Given that text (the word "text" is used to refer to both verbal only and multimedia text) is a primary means of acquiring and communicating knowledge, development of effective reading skills is crucial to the success of any society in a global knowledge-based economy. The research being reported led to the design of a Web-based reading environment that fundamentally extends the ways in which readers interact with text. This research had two purposes: (a) to design and develop a Web-based reading environment that supports the use of a set of reading strategies, and (b) to investigate the impact of this Web-based reading environment on readers' memory and understanding of an instructional unit on the human heart.

## Introduction

Students, particularly those in Science, Technology, Engineering, and Mathematics (STEM) fields, are faced with a situation in which they are expected to develop mastery of substantial and ever-growing bodies of knowledge in a limited time (Yore, Bisanz, & Hand, 2003). This requires adeptness with text which unfortunately is lacking in many US students. A report by the RAND Reading Study Group (Snow, 2002) noted that US students' lack of proficiency in reading puts them at a clear disadvantage with regard to learning from informational texts when compared with their peers in other countries. In fact, 11th grade students in the US were placed close to the bottom of all industrialized countries in a comparison on reading achievement. This deficiency in reading skills leads many students to drop out of STEM fields. There is also substantial evidence of an achievement gap in reading between low-economic status, at-risk students and their more privileged peers (NCES, 2000). At the same time, the RAND Reading Study group (Snow, 2002) noted that our knowledge of cognitive processes and strategies related to reading comprehension is inadequate to sufficiently reform comprehension instruction. This state of affairs has motivated work on a Web-based reading environment that is designed to extend the ways in which readers engage with informational text (Kidwai, 2009).

### Reading Comprehension

Reading for understanding is a non-trivial task, particularly when the reader is unfamiliar with the material being read. Even if readers can understand individual words in a text, they may not be able to link meanings of these words so as to be able to derive an understanding of larger chunks of text (Kintsch, 1988; van Dijk & Kintsch, 1983). Comprehension, defined as the ability to understand the ideas conveyed in a text, involves the creation of a mental representation of the text being read. The representation that is generated is a function of the information contained in the text and the reader's prior knowledge, which includes the reader's knowledge of the subject matter as well as the strategies for comprehension that the reader applies on the text (Mannes & George, 1996).

The quality of the reader's memory and understanding of a text is a function of the quality of the mental knowledge representations that are generated upon reading the text (Kintsch, 1988; van Oostendorp & Goldman, 1999). Higher quality knowledge representations directly impact an individual's ability to apply knowledge in novel situations. For example, the quality of knowledge representations is a known indicator of the probability of success in problem-solving situations (McNamara, E. Kintsch, Songer, & W. Kintsch, 1996).

### The Technology of Text

Humankind has witnessed revolutionary advances in technologies in the past centuries, yet the "technology of text" (the form in which text is presented to readers) has remained more or less unchanged (with few exceptions such as hyperlinks and animation) since the time of the handwritten scrolls and the invention of the printing press. Research on comprehension processes demonstrates a dissonance between the form in which text is presented to readers and the cognitive processes that successful readers engage in when they read (DeStefano & LeFevre, 2007; Bannert, 2004; Gervais, 2007, Zumbach & Mohraz, 2008). For example, instead of attempting to understand a large amount of text all at once, skilled readers use a "divide and conquer" strategy which leads them to identify and tackle smaller sections of text. Only when they are comfortable with their understanding of smaller sections of the text do these readers begin to integrate their understanding of the smaller pieces into a larger whole. Even though successful readers routinely engage in such bottom-up processes, the current technology of text does not allow readers to "manipulate" text in ways that support such processing. That is, even though readers mentally break up, reorganize, and reassemble sections of text that they read, they are not able to manipulate printed text, on paper or online, in ways that reflect these operations.

## Design of Web-based Reading Environment

The research being reported led to the design of a Web-based reading environment that fundamentally extends the technology of text by allowing readers to manipulate text in unique ways (Kidwai, 2009, see also AlAgha & Burd, 2009; Triantafillou, Pomportsis, & Demetriadis, 2003; Wolf, 2003). The design of the reading environment was informed by Kintsch's Construction-Integration (CI) model (Kintsch, 1988; van Dijk & Kintsch, 1983); Paivio's dual coding theory (Sadoski & Paivio, 2004; see also Ainsworth, 1999); research on the role of metacognitive processes such as comprehension monitoring (Palincsar & Brown, 1984; Pressley & Ghatala, 1990; Pressley & Harris, 2006), note-taking (Kiewra, DuBois, Christian, McShane, Meyerhoffer, & Roskelley, 1991) and summarization strategies (Armbruster, Anderson, & Ostertag, 1987; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989). Theories in human-computer interaction (Carroll, 2003) informed the design of the user-interface of the Web-based reading environment. In particular, five reading strategies were designed in the reading environment:

(a) *Text-macrostructure (chunking) strategy*: A reader can chunk the instructional unit on the human heart into smaller sections by moving elements that make up the instructional unit into *tabs* that they create on the interface of the Web-based learning environment. A reader can also label these tabs. In this way each tab on the user interface corresponds to a reader-generated *section* of instructional unit. An example of a sequence of steps that a reader would take in executing this strategy is provided below (see Figure 1, Figure 2, Figure 3). It is expected that in the process of reorganizing the instructional unit at the macro-level, the reader would begin to develop a macro-level understanding of the text—the "big picture." Being able to *customize* the instructional unit in this way should also motivate the reader to engage with it deeply.

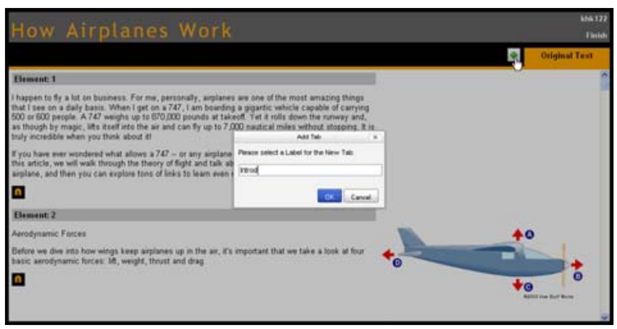



*Figure 1*. Text-macrostructure strategy: Creating a new section (tab on the interface).

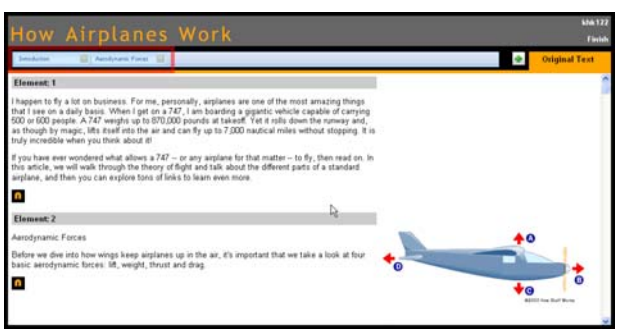



*Figure 2*. Text-macrostructure strategy: Two new sections are created (top-left hand corner).

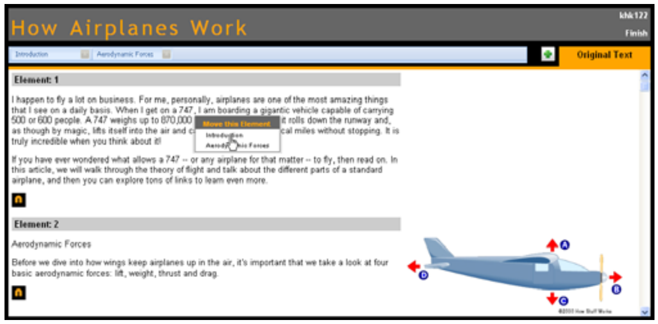



*Figure 3*. Text-macrostructure strategy: An element is moved to a section (tab).

b. *Summarization strategy*: A reader can write summaries for each section (tab) that she creates (see Figure 4). It is expected that in the process of writing summaries the reader would access her *macro-level understanding* of the instructional unit. In the process of writing summaries, gaps in macro-level understanding, if any, would be exposed. In the event the reader found gaps in understanding, she could take action—for example, rereading a particular section of the instructional unit (Armbruster, Anderson, & Ostertag, 1987).

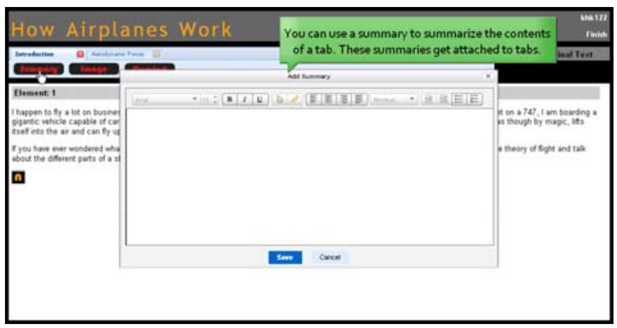



*Figure 4*. Summarization strategy.

c. *Imagery strategy*: A reader can select an image from a library of images that is most representative of the contents of a given section (see Figure 5). This strategy is expected to encourage readers to integrate knowledge across modalities, which in turn should lead to creation of well integrated and robust knowledge representations (Sadoski & Paivio, 2004; see also Ainsworth, 1999).

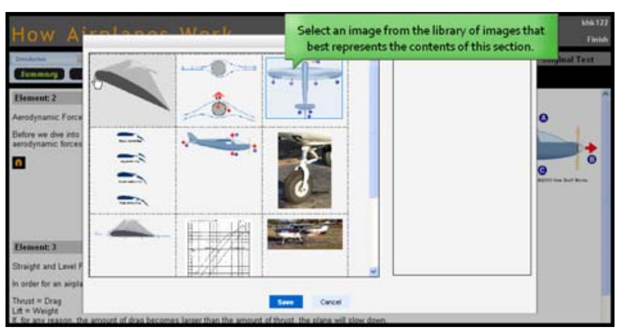



*Figure 5*. Imagery strategy.

d. *Reading self-assessment (comfort-meter) strategy*: A reader can mark on a scale of 1-5 how comfortable she feels with her understanding of each section of the instructional unit (see Figure 6). It is expected that a metacognitively aware reader would realize that not all sections of the instructional unit are equally challenging. On the basis of this knowledge, the reader could decide to allocate cognitive resources to the more difficult sections, for example by choosing to revisit and reread them (Pressley & Harris, 2006; Pressley & Ghatala, 1990).

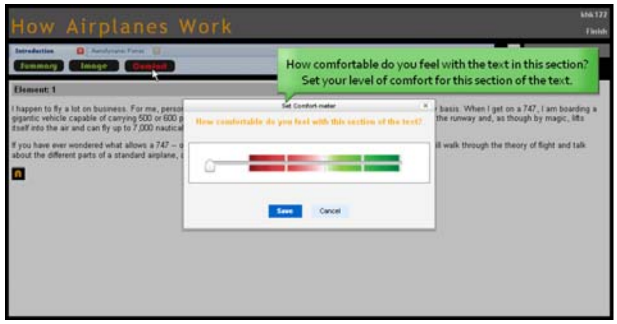



*Figure 6*. Reading self-assessment (comfort-meter) strategy.

e. *Note-taking strategy*: A reader can take notes on an *element-by-element* basis—these notes are attached to the elements that make up the instructional text (see Figure 7). It is expected that note-taking will support comprehension by allowing the reader to externalize her understanding, maintain attention, and provide a summary of the main points for a given element in the instructional unit (DiVesta & Gray, 1972; Peper & Mayer, 1986; see also Kiewra et al., 1991).

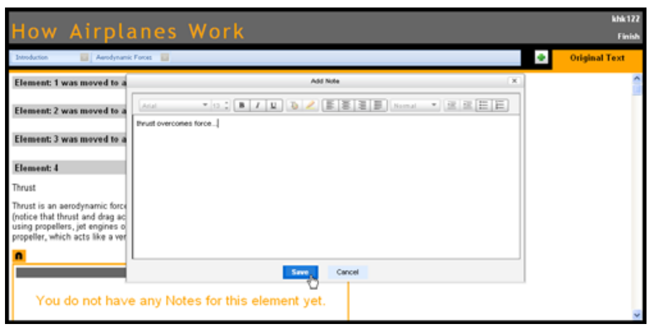



*Figure 7*. Note-taking strategy.

## Research Goals

This research had two purposes: (a) to design and develop a Web-based reading environment that supports the use of a set of reading strategies, and (b) to investigate the impact of this Web-based reading environment on readers' memory and understanding of an instructional unit on the human heart. These purposes led to four major research goals (also see Table 1):

1. Guide the design of a Web-based reading environment that supports the use of a set of five reading strategies.
2. Establish validity and reliability of Web-based measurement instruments that were designed to measure memory and understanding of the instructional unit on the human heart at four levels of knowledge representation—micro-textbase, macro-textbase, micro-situation model and macro-situation model.
3. On the basis of these measurement instruments, determine whether participants who read the instructional unit on the human heart in the Web-based reading environment developed better memory and understanding of the instructional unit when compared with participants who read the same text on a regular Web page.
4. Develop an understanding of participants' use of the five reading strategies in the Web-based reading environment and their experience in the Web-based reading environment.

Table 1. Research goal(s), number of participants, and materials in the four phases of the design-based research investigation

| Phase | Research Goals | Number of Participants | Primary Materials |
|---|---|---|---|
| Phase I | Research Goal 1: Guide the design of a Web-based learning environment. | 10 | Paper prototype of the Web-based learning environment with only the text-macrostructure strategy |

| | | | |
|---|---|---|---|
| Phase II | Research Goal 2: Establish validity and reliability of Web-based instruments that are designed to measure memory and understanding of the instructional unit on the human heart at four levels of knowledge representation—micro-textbase, macro-textbase, micro-situation model and macro-situation model. | 57 | Web-based human heart tests; instructional unit on the human heart presented on a regular Web-page |
| Phase III | Research Goal 1: Guide the design of a Web-based learning environment.<br><br>Research Goal 4: Develop an understanding of readers' use of the five reading strategies in the Web-based learning environment and their experience in the Web-based learning environment. | 43 | Web-based and paper-form of the Web-based learning environment with all five reading strategies; animated demonstration of the Web-based learning environment; Web-based human heart tests; feedback survey |
| Phase IV | Research Goal 3: Determine the impact of reading in the Web-based learning environment on readers' memory and understanding of the instructional text on the human heart.<br><br>Research Goal 4: Develop an understanding of readers' use of the five reading strategies in the Web-based learning environment and their experience in the Web-based learning environment. | 685 | Web-based learning environment with all five reading strategies; two animated demonstrations of the Web-based learning environment; Web-based human heart tests; feedback survey |

Participants in the study were undergraduate students at a large public university in the northeastern United States. Potential benefits of using the strategies in the reading environment include: (a) Cognitive: Strategies such as the text-macrostructure strategy will allow readers to externalize the mental representations that they create as they read and process text. These external representations of the text should free up cognitive resources that would otherwise be used to store and recall these representations from long term memory (Niederhauser, Reynolds, Salmen & Skolmoski, 2000; Paas, Renkl & Sweller, 2004; Paas, Tuovinen, Tabbers & Van Gerven, 2003; Mayer & Moreno, 2003). Cognitive resources that are freed-up could be allocated to comprehension processes. (b) Metacognitive: Writing summaries and setting the comfort-meter will give readers an opportunity to evaluate their understanding of a given section of the instructional text (Schwartz, Andersen, Hong, Howard & McGee, 2004). Metacognitively aware readers could use this opportunity to find gaps in their understanding and take remedial action. (c) Motivational: The ability to customize text can be motivating (Lawless & Brown, 1997; Wigfield, & Guthrie). Strategies such as text-macrostructure, note-taking, summarization, imagery, and reading self-assessment, will allow readers to customize the text to suit their learning style and their level of comfort with particular sections of the instructional text.

## Results

A design-based research methodology guided a series of studies that addressed the research goals in the study. These studies were conducted in four phases. Findings from each of the first three phases of the investigation were used to inform and update the design of the Web-based reading environment, the Web-based measurement instruments, and the research design of the subsequent phase. The investigation culminated in a large-scale quasi-experimental study.

In the Phase I study, participants' use of the text-macrostructure or chunking strategy was investigated with the help of an early-stage paper prototype of the Web-based reading environment. Participants' organization of the individual *elements* that made up the instructional unit on the human heart into *sections,* and the labels that participants gave these sections were markedly *different* across the six participants who completed this study. This evidence indicated that participants in the study made use of the text-macrostructure strategy to develop *unique macro-level representations* of the instructional unit on the human heart. Findings from this study provided support for incorporating the text-macrostructure strategy in the Web-based reading environment. This strategy played a key role in the overall design of the Web-based reading environment.

In the Phase II study, Web-based measurement instruments that measured participants' knowledge of the instructional unit on the human heart at four levels of knowledge representation—micro-textbase, macro-textbase, micro-situation model, and macro-situation model—were designed and validated. Technology used to develop the measurement instruments (implementation of the autocomplete design-pattern, drag-and-drop, and AJAX) and online delivery of measurement instruments was tested. The Web-based measurement instruments developed in this study were used in the subsequent phases of the investigation.

One-on-one sessions with 43 participants were conducted in the Phase III study. Participants were found to engage deeply with the reading task in the Web-based reading environment spending nearly twice as much time reading when compared with participants in the Phase II study who read the instructional unit on the human heart on a regular Web page. In preparation for the Phase IV study, an animated demonstration that provided participants an overview of the user interface of the Web-based reading environment was prepared and tested. Feedback questions related to participants' experience in the Web-based reading environment and the Web-based measurement instruments were prepared and tested. On the basis of data collected from interviews with participants, responses to the feedback survey, think-aloud and video data, numerous updates were made to the user interface and the design of the reading strategies in the Web-based reading environment.

On the basis of observations in the Phase III studies, a decision was made to develop two variants of the animated demonstration for the Web-based reading environment. The first demonstration was similar to the demonstration used in the Phase III studies. The objective of this demonstration was to provide participants an orientation of the user interface of the Web-based reading environment. In addition to providing participants an orientation of the user interface of the Web-based reading environment, the second demonstration provided *explicit instruction* on the five reading strategies.

The Phase IV study investigated the impact of the Web-based reading environment on readers' memory and understanding of an instructional unit on the human heart. Two levels of the independent variable—the Web-based reading environment—were set up: (a) Web-based reading environment with demonstration that *did not include* explicit instruction on the reading strategies (T1); and (b) Web-based reading environment with demonstration that *included* explicit instruction on the reading strategies (T2). In the control condition participants read the instructional unit on the human heart on a regular Web page. The six dependent variables in the study included scores on four tests that measured participants' memory and understanding of the instructional unit on the human heart (micro-textbase, macro-textbase, micro-situation model, and macro-situation model), time spent reading the instructional unit on the human heart, and the time spent responding to test questions. Six research questions were pursued: (a) Did the Web-based reading environment help participants develop better *memory* (micro-textbase and macro-textbase) of the instructional unit on the human heart, and was there an effect of the explicit instruction on reading strategies? (b) Did the Web-based reading environment help participants develop better *understanding* (micro-situation model and macro-situation model) of the instructional unit on the human heart, and was there an effect of the explicit instruction on reading strategies? (c) Did participants in the three experimental conditions spend equal amount of time reading the instructional unit on the human heart? (d) Did participants in the three experimental conditions spend equal amount of time responding to Web-based human heart tests? (e) How did participants use the five reading strategies in the Web-based reading environment and what was their experience in the Web-based reading environment? (f) What was participants' experience with the Web-based human heart tests?

Participants in the Web-based reading environment group with demonstration that included explicit instructions on the reading strategies (T2) performed significantly better ($p < .10$) than participants in the control group on the macro-textbase, micro-situation model, and macro-situation model tests ($p = .076$, $p = .079$, and $p = .012$ respectively). There was no difference in performance on the micro-textbase test ($p = .274$). Performance of participants in the Web-based reading environment group with demonstration that did not include explicit instructions on the reading strategies (T1) was unexpected; they scored lower than participants in the control group on all four tests.

Participants in the two Web-based reading environment conditions (T1 and T2) spent nearly twice as much time reading the instructional unit on the human heart when compared with participants in the control group who read the instructional unit on the human heart on a regular Web page ($p = .000$ and $p = .000$ respectively).

Participants in the Web-based reading environment group with demonstration that did not include explicit instructions on the reading strategies (T1) spent significantly less time responding to questions on the

Web-based human heart tests when compared with participants in the control group and the Web-based reading environment group with demonstration that included explicit instructions on the reading strategies (T2) ($p$ = .001 and $p$ = .001 respectively).

Feedback from participants in the Web-based reading environment groups indicated that they engaged deeply with the instructional text in the Web-based reading environment; many of the participants were able to use the five reading strategies effectively, several participants thought that they benefited from using these reading strategies. Overall, 87% of the participants noted that they had a favorable experience in the Web-based reading environment. There was a significant difference between the Web-based reading environment group with demonstration that did not include explicit instructions on the reading strategies (T1) and the Web-based reading environment group with demonstration that included explicit instructions on the reading strategies (T2) [$\chi^2$ (1, $N$ = 353) = 8.10, $p$ = .004]; 93% of the participants in Treatment 2 (T2) voted favorably, as compared to 83% of the participants in Treatment 1 (T1). Furthermore, 89% of the participants in Treatment 2 (T2) noted that they would have liked to use the Web-based reading environment in the following semester. In comparison, this percentage was only 75% for participants in Treatment 1 (T1). This difference was statistically significant [$\chi^2$ (1, $N$ = 351) = 13.88, $p$ = .000].

Overall, 82% of the participants found the experience of responding to questions on the Web-based human heart tests to be favorable. There was a significant difference between the three experimental groups [$\chi^2$ (2, $N$ = 545) = 16.62, $p$ = .000]. Participants in the control group and the Web-based reading environment group with demonstration that included explicit instructions on the reading strategies (T2) rated their experience to be more positive (89%, 84% respectively) than participants in the Web-based reading environment group with demonstration that did not include explicit instructions on the reading strategies (T1) (73%).

## Conclusion

Reading comprehension is contingent upon the successful execution of a series of cognitive and metacognitive processes. The reading strategies in the Web-based learning environment support a subset of these comprehension processes. Findings from this research stand to inform our understanding of these comprehension processes, and ways in which affordances of current Web technologies can be used to design compelling reading environments and assessments that measure learning that occurs in these environments. The Web-based learning environment also has potential for impacting the practice of reading strategies instruction. According to one participant in the study, "*Yes, I think it [reading environment] not only allowed me to learn a few things about the heart, but it also taught me that I work better when breaking text apart and categorizing the material. It also allows you to actually attach a meaning to the text instead of just reading it*."

## References


Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33 131-152.

Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly, 22*(3), 331-346.

Barab, S. (2006). Design-based research: A methodological toolkit for the learning scientist. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 153-169). Cambridge, United Kingdom: Cambridge University Press.

Bell, P., Hoadley, C. M., Linn, M. C. (2004). Design-based research in education. In M. C. Linn, E. A. Davis, & P. L. Bell (Eds.), *Internet environments for science education* (pp. 73-85). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Carroll, J. M. (2003). *HCI models, theories, and frameworks: Toward a multidisciplinary science*. San Francisco, CA: Morgan Kaufmann Publishers.

Collins, A., Joseph, D., & Bielaczyc, K. J. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15-42.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Dwyer, F. M., & Lamberski, R. J. (1977). *The human heart: Parts of the heart, circulation of the blood and cycle of blood pressure.* State College, PA: Learning Services.

Graesser, A. C. (2007). An introduction to strategic reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 3-26). New York: Lawrence Erlbaum Associates, Publishers.

Graesser, A. C., Millis, K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*(1), 163-189.
Hoadley, C. (2002). Creating context: Design-based research in creating and understanding CSCL. In G. Stahl (Ed.), *Computer Support for Collaborative Learning* 2002 (pp. 453-462). Mahwah, NJ: Lawrence Erlbaum Associates.
Hoadley, C. (2004). Methodological alignment in design-based research. *Educational Psychologist, 39*(4), 203-212.
Hoadley, C., Hsi, S., & Berman, B. P. (1995). The Multimedia Forum Kiosk and SpeakEasy. In P. Zellweger (Ed.), *Proceedings of the third ACM international conference on multimedia* (pp. 363-364). San Francisco, CA: ACM Press.
Hoadley, C., Linn, M. C., Mann, L. M., & Clancy, M. J. (1996). When and why do novice programmers reuse code? In W. Gray & D. Boehm-Davis (Eds.), *Empirical Studies of Programmers, Sixth Workshop* (pp. 109-130). Norwood, NJ: Ablex Publishing Company.
Hoadley, C., & Pea, R. (2002). Finding the ties that bind: Tools in support of a knowledge-building community. In K. A. Renninger & W. Shumar (Eds.), *Building virtual communities: Learning and change in cyberspace* (pp. 321-354). New York: Cambridge University Press.
Kidwai, K. (2009). *A Design-Based Research Investigation of a Web-Based Learning Environment Designed to Support the Reading Process* (Doctoral dissertation). The Pennsylvania State University, University Park, PA.
Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95*(2), 163-182.
Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.
McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43.
Mannes, S. M., & George, M. S. (1996). Effects of prior-knowledge on text comprehension: A simple modeling approach. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 115-139). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
Mayer, R. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, 43-52.
Meyer, B. J. F., & Wijekumar, K. (2007). A Web-based tutoring system for the structure strategy: Theoretical background, design, findings. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 347-374). New York: Lawrence Erlbaum Associates, Publishers.
Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117-175.
Pressley, M., & Ghatala, E. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist, 25*(1), 19-33.
Pressley, M., & Harris, K. (2006). Cognitive strategies instruction: From basic research to classroom instruction. In P. A. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 265-286). New York: MacMillan.
Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J. A., & Kurita, J. A. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal, 90*(1), 3-32.
Sadoski, M., & Paivio, A. (2004). A dual-coding theoretical model of reading. In R. B. Ruddell & N. J. Unrao (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1329-1362). Newark, DE: International Reading Association.
Schwartz, N. H.. Andersen, C., Hong, N., Howard, B., & McGee, S. (2004). The influence of metacognitive skills on learner's memory of information in a hypermedia environment. *Journal of Educational Computing Research*, 31, 77-93.
Snow, C.E. (2002). Reading for understanding: Toward an R&D program in reading comprehension. Santa Monica, CA: RAND Education.
van Dijk, T., & Kintsch, E. (1983). *Strategies of discourse comprehension*. New York: Academic.
van Oostendorp, H., & Goldman, S. R. (Eds.). (1999). *The construction of mental representations during reading*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
Yore, L., Bisanz, G.L., & Hand, B.M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education, 25*, 689-725.
Zwaan, R. A., & Radvansky. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.

# Helping Students Make Controlled Experiments More Informative

Kevin W. McElhaney, Marcia C. Linn
University of California, Berkeley, 4523 Tolman Hall MC1670, Berkeley, CA 94720, USA
kevin777@berkeley.edu, mclinn@berkeley.edu

**Abstract:** We examine how encouraging students to *compare* rather than *isolate* variables affects their experimentation strategies and insights. We designed a week-long, technology-enhanced inquiry module on car collisions that logs students' interactions with a visualization. Physics students (N=166) were assigned to conditions that prompted them either to isolate or compare variables. Students responded to pretests, posttests, and embedded prompts that assessed students' understanding of motion graphs and collisions. Both groups made significant pretest to posttest gains. Students in the *compare* treatment used more diverse experimentation strategies than students in the *isolate* treatment. *Compare* students made nuanced interpretations of collision events based on threshold values. Case studies illustrate how comparing rather than isolating helped students use wide-ranging strategies to reach complex insights. The findings illustrate the value of encouraging multiple approaches to experimentation and connecting experimentation to real-life contexts.

## Introduction

This study examines how two different types of experimentation goals led students toward different investigation strategies and insights about a complex science problem. We designed *Airbags: Too Fast, Too Furious?*, a week-long inquiry module for high school physics classes that guides students through an investigation about car collisions. In *Airbags*, students conduct experiments using a visualization to investigate what factors increase drivers' risk for injury from an airbag. The software logs students' experimentation choices and explanations to provide detailed information about their inquiry activities and inferences. This study examines two main research questions: (1) What is the overall impact of *Airbags* on students' ability to interpret and construct motion graphs? and (2) How does prompting students to *compare*, rather than *isolate,* variables in their experiments guide students toward different experimentation strategies and insights?

## Rationale

Early research on scientific reasoning addressed children's ability to isolate variables in knowledge-lean experimentation contexts. For instance, Inhelder and Piaget (1958) designed a task [later adapted by Kuhn and Phelps (1982)] that asked subjects to determine what combination of colorless fluids would yield a specific reaction outcome. Siegler and Liebert (1975) examined the ways subjects determined how an electric train runs on the basis of four binary switches (though in actuality, a researcher operated the train using a secret switch to ensure that subjects would test all 16 combinations). These studies examined experimentation as domain-general logical inference, as subjects had no information on which to base testable hypotheses. In these situations, subjects could make valid inferences only by isolating variables to logically eliminate possibilities.

Over time, research has increasingly examined knowledge-rich contexts and revealed the important role of context-specific knowledge in experimentation. For example, studies show that children are more likely to test plausible rather than implausible hypotheses (Klahr, Fay, and Dunbar, 1993), focus on variables they believe to be causal (Kanari & Millar, 2004), and use experiments to achieve specific outcomes rather than test hypotheses (Schauble, 1996). Though learners' ideas about the investigation context may lead them toward invalid experimental designs or inferences, students may also use ideas productively, such as by narrowing the range of testable values or eliminating implausible explanations. Tschirgi (1980) argued that children's tendency to use "invalid" strategies when determining the ingredients needed to bake a good cake is reasonable, given real-life goals of reproducing positive results (good cakes) and eliminating negative ones (bad cakes). Koslowski (1996) also argued that using prior knowledge to generate and interpret evidence is a good strategy, particularly when understanding mechanisms informs the interpretation of outcomes.

In *Airbags*, students rely on many types of context-specific ideas to draw conclusions. Students' everyday understanding of motion, their physics domain knowledge about motion graphs, and evidence from the World-Wide Web all contribute to the way students design and interpret their experiments. Our previous work on *Airbags* (McElhaney & Linn, 2008) compared students who were constrained to conduct a specific number of trials to unconstrained students. During planning, constrained students attended more to the logistics of isolating variables, resulting in more controlled trials than unconstrained students. Unconstrained students, who attended more to the relationships among collision factors, variables, and outcomes, demonstrated a superior understanding of the situation. This result suggests that incorporating a context-specific understanding into experimentation was more important than employing "valid" strategies.

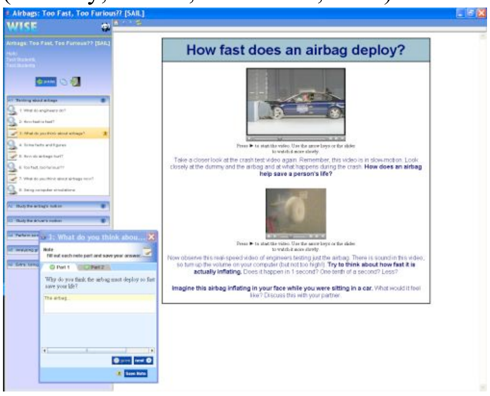
The current study extends our prior work by examining how experimentation goals can highlight the nature of the variables and make students' experimentation more informative. Research shows that goals for experimentation can influence learners' strategies and inferences, such as when children use different strategies to achieve positive outcomes than negative ones (Tschirgi, 1980) or when learners use different approaches for optimization than for rule-generation (Schauble, Klopfer, & Raghavan, 1991). This study draws from research in science instruction showing that comparisons can highlight key features of variables (Clement, 1983; Linn, 2005; Schwartz & Bransford, 1998) and investigates whether goals that encourage students to *compare* variables promote different insights about *Airbags* than goals that encourage students to *isolate* variables.

This study also explores how students come to understand *thresholds* (values of a particular variable above or below which the outcome is independent of the other variables). The presence of thresholds requires students to consider factors other than covariation to achieve a complete understanding of *Airbags*. Though students may readily observe the effects of thresholds by isolating variables, we hypothesized that comparing variables would highlight the distinct characteristics of each variable, helping sophisticated learners achieve a more nuanced understanding of *Airbags* based on thresholds as well as covariation. We also expected that the more familiar and tractable task of isolating variables would be more beneficial for less sophisticated learners.

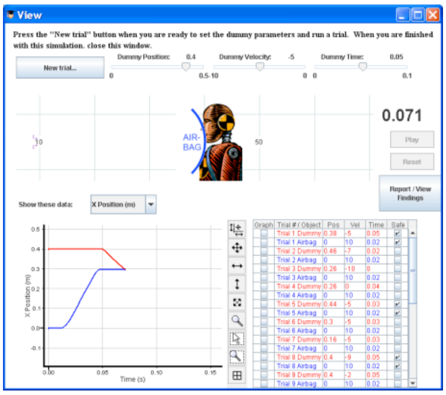
## Methods

### Module and visualization design

We designed *Airbags* (Figure 1a), using the Web-based Inquiry Science Environment (Linn, Davis, & Bell, 2004). *Airbags*' learning goals are (1) how motion graphs represent one-dimensional motion and (2) what factors make airbags likely to injure drivers. Students experiment using a dynamic visualization (Figure 1b) designed by the Concord Consortium (www.concord.org). Students investigate the role of three *collision factors* on the driver's safety by experimenting with three *motion variables*. Table 1 summarizes factors, variables and questions for each treatment condition. To conduct a trial, students select an investigation question (or indicate that they are *Just exploring*), specify the variable values, and run the simulation. For each trial, students judge whether the driver was "safe" (encountered a completely inflated airbag) or "unsafe" (encountered the airbag within its deployment zone). The visualization presents the motion of the airbag and driver using graphs that coordinate with an animation. The visualization also helps students manage their trial history, sort trial outcomes, and compare multiple trials. The software logs students' experimentation choices for subsequent analysis (Buckley, Gobert, & Horwitz, 2006).




(a)




(b)

Figure 1. (a) The first activity of *Airbags*. (b) The experimentation visualization.

Students may use two types of inferences to explain their findings: covariation and threshold values. First, over a particular range of values, each of the three variables covaries with the time that elapses before the driver and airbag collide. Tall drivers, low speed collisions, and a large crumple zone therefore make a driver more likely to encounter a fully inflated airbag than short drivers, high speed collisions, and a small crumple zone. Second, two threshold values (for position and time) determine situations where the likelihood of injury is invariant: (1) short drivers who sit within an airbag's zone of deployment will *never* encounter a fully inflated airbag, and (2) for sufficiently tall drivers, if the duration of the crumple zone exceeds the deployment time for the airbag, drivers will *always* encounter a fully inflated airbag. Responses to embedded assessments indicate whether students attribute their findings to covariation and/or thresholds.

### Study design

Groups in the *isolate* and *compare* conditions received investigation questions that encouraged them to isolate or compare variables, respectively. Except for the investigation questions, the modules used for each condition were identical. We used a pretest/posttest experimental design with embedded assessments and two comparison conditions. We randomly assigned student groups to one of the conditions using a stratified approach to distribute their ability equally across the two conditions. To keep instruction similar across conditions, teachers and researchers focused class discussions on graph interpretations rather than on investigation questions and refrained from guiding individual students on designing experiments.

Table 1: Collision factors, variables, and investigation questions for the *isolate* and *compare* conditions.

| Collision factor | Motion variable | Investigation question | |
|---|---|---|---|
| | | *Isolate* condition | *Compare* condition |
| Driver height | Position | Are TALL or SHORT drivers more likely to be injured by a deploying airbag? | Does the DRIVER'S HEIGHT make the biggest difference in whether the driver is injured? |
| Collision speed | Velocity | Do HIGH or LOW SPEED collisions make drivers more likely to be injured by a deploying airbag? | Does the COLLISION SPEED make the biggest difference in whether the driver is injured? |
| Car crumpling | Time | Does MORE CRUMPLING or LESS CRUMPLING make drivers more likely to be injured by a deploying airbag? | Does HOW MUCH THE CAR CRUMPLES make the biggest difference in whether the driver is injured? |

## Participants

Physics students (N=166) at five socially diverse high schools in the United States studied *Airbags*. Most students worked in dyads on the module (unpaired students worked alone). Students were usually grouped with other students of nearly equal ability. Most teachers had taught the topic using previous versions of *Airbags*.

## Data sources and scoring

### Pretests and posttests

Ten constructed response pretest and posttest items assessed students' ability to interpret and construct graphs of one-dimensional motion in a context outside of *Airbags*. We administered pretests individually to students the day before the beginning of *Airbags* and posttests within a few days of completion. We scored pretest and posttest responses from zero to four using knowledge integration rubrics (Linn, Lee, Tinker, Husic, & Chiu, 2006) that measured how well students connected characteristics of motion to the features of motion graphs.

### Embedded assessments

*Use of the Control-of-Variables Strategy (CVS).* We used students' experimentation sequences (as captured online in log files) to compute a CVS score, the percentage of each group's trials that were part of a controlled comparison between successive trials using the variable appropriate to the chosen investigation question. We used only trials that specified one of the three investigation questions (i.e. not *Just exploring*) for this score. This proportion score indicates the extent to which students mindfully used CVS to investigate the three questions.

*Interpreting and constructing Airbags graphs.* Twelve items asked students to interpret motion graphs from the visualization or generate graphs that represented a collision situation. Each group received a COLLISION GRAPHS score that captured how well they connected characteristics of the graphs to the events in a collision. We scored these explanations from zero to five using knowledge integration rubrics.

*Understanding of covariation and thresholds.* Three items prompted students to explain their answers to the three investigation questions listed in Table 1. We used these responses to determine whether each group attained a covariation-based and/or thresholds-based understanding of the *Airbags* situation. To examine how frequently students attributed their findings to covariation, we counted the number of responses by each group that described covariation between the factor or variable in question and elapsed time before the driver encounters the airbag. To examine how frequently students attributed their findings to thresholds, we counted the number of responses by each group that used at least one of the thresholds in support of their finding.

### Videorecords

We videorecorded 12 dyads as they engaged in the experimentation activity, so that we could closely examine the discussions that occurred during experimentation within specific dyads. We chose two of these dyads for a case comparison to illustrate how the investigation questions they received influenced their experimentation strategies and insights. A more detailed description of these two dyads and the reasons we chose them for analysis follows the presentation of the results of the comparison study.

# Results

## Impact of *Airbags* on students' understanding of motion graphs

Students made moderate, significant pretest-posttest gains [M = 29.99, SD = 6.66 (pre); M = 32.76, SD = 5.96 (post), $t(128) = 5.17$, $p < .001$ (two-tailed), $d = 0.44$]. Gains were positive for all five schools, and significant for four of the schools. Low prior knowledge learners made the greatest gains. The gains indicate that *Airbags* was successful in helping diverse learners interpret and construct motion graphs. The gains are impressive considering that all students had completed a kinematics unit shortly before studying *Airbags*, and they represent value added to a traditionally taught kinematics unit.

The difference in posttest scores between the two conditions was not significant. We did not expect differences because the experimentation activity generally comprised just 10-20% of the total time spent on *Airbags*. Students in both conditions therefore had many opportunities to improve their graphing knowledge other than by experimentation. Furthermore, there were no significant differences between the conditions on the COLLISION GRAPHS scores, suggesting that the *isolate* and *compare* prompts were equally effective in helping students connect the characteristics of the *Airbags* graphs to the collision events.

To examine subtle impacts of experimentation goals on students' strategies and insights, we sorted student groups into low, middle, and high prior knowledge tertiles using the mean pretest score for each group. Our exploratory analyses show variation in students' strategies and inferences by tertile.

## Impacts of prior knowledge on students' strategies and inferences

CVS scores showed that the high tertile groups conducted a significantly higher proportion of controlled trials than the low and middle tertile groups [M = .79, SD = .30 (high), M = .45, SD = .35 (low/middle), t(79) = 4.39, $p < .001$ (two-tailed)]. This difference suggests that high prior knowledge students as a whole were more focused on controlling variables in their investigations.

Overall, about 30% of the students generated covariation-based explanations of the collision events. The percentage was somewhat higher for the high (38%) and middle (36%) tertile groups than for the low (19%) tertile groups, but this difference was not significant, indicating that a covariation-based understanding of *Airbags* was accessible to students at all prior knowledge levels. However, just 2% of low and middle tertile students generated thresholds-based explanations, compared to 31% of the high tertile students. A Wilcoxon rank-sum test showed that the difference in the average number of thresholds-based explanations was significant [M = .59, SD = .98 (high), M = .038, SD = .28 (low/middle), U = 3.78, $p < .001$]. This result suggests that only sophisticated learners were able to achieve a thresholds-based understanding of *Airbags*.

## Impacts of experimentation goals on students' strategies and inferences

CVS scores showed significant differences between the conditions only for the high tertile students [M = .94, SD = .06 (*isolate*), M = .70, SD = .35 (*compare*), t(27) = 2.22, $p = .035$ (two-tailed)]. A close examination of these data revealed that while virtually all the *isolate* groups devoted nearly all their trials to controlled comparisons between successive trials, one-third of the *compare* groups used at least half their trials for other strategies. (The case comparison that follows will illustrate some of these strategies.) This result suggests that though most students used similar approaches to investigate the *isolate* and *compare* questions, the *compare* questions led some high prior knowledge students to use alternative strategies to CVS in their investigations.

There were no differences in the number of covariation-based explanations between the two conditions, indicating both conditions were equally effective in leading students toward a covariation-based understanding of *Airbags*. However, we observed significant differences in the number of thresholds-based explanations for the high tertile students. Just 8% of the high tertile *isolate* groups generated thresholds-based explanations, compared to 44% of the *compare* groups. A Wilcoxon rank-sum test showed that the difference in the average number of thresholds-based explanations for high tertile students was significant [M = .091, SD = .30 (*isolate*), M = .89, SD = 1.13 (*compare*), U = 2.09, $p = .037$]. This finding suggests that the *compare* questions promoted a thresholds-based understanding of *Airbags*, though this difference occurred only with sophisticated learners.

# Case comparison

We use a case comparison to illustrate some ways the *isolate* and *compare* tasks might have led sophisticated students toward different experiences with the *Airbags* visualization. We chose two similar high-tertile dyads for a case comparison. The students in these dyads were enrolled in the same honors physics curriculum with the same teacher and had similar pretest and posttest scores. Furthermore, all four students were concurrently enrolled in calculus and thus had strong mathematics skills. Because the student population and array of schools used for this study was highly diverse, no single case can represent the "typical" experience students have with *Airbags*. These cases rather aim to illustrate how the *compare* questions might have prompted sophisticated students to consider aspects of the *Airbags* situation that the more traditional *isolate* question might not have.

## Case 1: Brett and Eric (*isolate* condition)

### Overview

Brett and Eric (pseudonyms) studied the *isolate* version of Airbags. They conducted 10 total trials. For their first four trials, they chose *Just exploring* as their goal, during which time they explored outcomes for default and extreme values of all three variables. They devoted the last six trials to conducting a pair of controlled trials for each of the three investigation questions. Their experimentation session lasted about 10 minutes. This analysis focuses on the three controlled trial pairs.

### Trials 5 & 6: "He's going to be safe, obviously"

In trials 5 and 6, Eric and Brett investigated whether tall or short drivers are more at risk. In trial 5, they examined the outcome for a short person. The following excerpt illustrates their variable choices for trial 6 and their interpretation of the two trials:

(84) E: …Short or tall. And now we have to move the guy back, cause he's taller. So we gotta keep everything except position. So move him back some. Like right there.
(85) B: He's going to be safe, obviously.
(86) E: He might not, let's just check. [They run trial 6.] Yeah. So mark that as safe. OK, put the graphs for the previous two. [They compare the graphs of trials 5 and 6]
(87) B: They're both safe.
(88) E: Yeah. So let's go to the next question.

Two things are apparent from this exchange. First, because trial 5 produced a "safe" outcome, Brett knew that the outcome of the trial 6 would also be "safe" before conducting it. However, rather than choosing a set of values that would provide them with more information, they simply ran the test. Second, in these two trials they achieved the same "safe" outcome, failing to provide strong evidence for the effect of the position variable on the risk to the driver. Further tests aiming to illustrate conditions that led to an "unsafe" outcome would have better informed their understanding. However, their variable choices and the brevity of their discussion about the results suggest they are focused more on isolating variables than on gaining insight about the situation.

### Trials 7 & 8: "We can only change one of them, we can't change multiple ones"

The following exchange occurred immediately after isolating the velocity variable in trials 7 and 8:

(93) E: …since we're doing, like, experiments, we can only change one of them, we can't change multiple ones.
(94) B: Yeah.
(95) E: Cause like in real life, there would be a combination of all three.

Eric's comments shed light on their commitment to using CVS throughout the activity. Eric believed that they were prohibited from using other strategies (though at no point does *Airbags* suggest how students should conduct their trials). Furthermore, the distinction he made between their experiments and "real life" indicates he believed these other strategies would be permissible in other contexts. Though Eric did not elaborate on what he meant by "a combination of all three", his words suggest that by isolating variables they aim to fulfill expectations imposed on them by the culture of classroom science.

### Trials 9 & 10: "Just go low first"

Brett and Eric used their final two trials to isolate the time variable. This exchange occurred as they decided how to conduct these two trials:

(97) E: So--more or less crumpling. So keep the velocity constant. So then, if there's less delay, that means more crumpling. So then, yeah let's do less crumpling.
(98) B: That's more crumpling.
(99) E: That's…yeah. So just go low first. Like there's low crumpling, almost no crumpling.
…
(103) E: OK, so then, that was low speed, so this is high speed. All right.
(104) B: More crumpling.
(105) E: Yeah, more crumpling.

To begin, Eric incorrectly stated the relationship between the crumpling factor and the time variable. Though this error could be conceptual, Eric's subsequent request to "just go low first" suggests that the correct relationship was unimportant to him. Because the precise nature of the relationship between the factor and the variable (direct or inverse) would not change how they employed CVS, the only decision Eric believed they need to make is whether to test the "low" or the "high" value first. Next, even after being reminded of the correct relationship, Eric attributed the time variable to the wrong factor entirely (speed, rather than crumpling). By this time, Eric was no longer attending to the nature of the variables and appears to have sequestered their experimentation strategy from their understanding of the *Airbags* context. At this point, the variables might as well have been X, Y, and Z rather than position, velocity, and time.

## Case 2: Joann and Linda (*compare* condition)

### Overview

Joann and Linda (pseudonyms) studied the *compare* version of *Airbags*. They began their experimentation by carrying out an initial plan to isolate each variable. They quickly abandoned that approach and employed other strategies such as testing extreme values and incrementally varying individual variables. Their experimentation session lasted about 30 minutes. This analysis will focus on three excerpts that illustrate the evolution of their investigation strategies and the insights they achieved by using these diverse approaches.

### Trials 1 - 4: Abandoning CVS

Joann and Linda used their first three trials to isolate the position variable and test its full range. As they decided on the values for the fourth trial, Linda began to reconsider their approach.

(41) L: I don't know. Maybe we just test ummm, like, test the position at, like, three different points. That's just so—that's just so many tests, never mind….

At first they were discouraged by the sheer length of their proposed approach, but after choosing some intermediate values for trial 4 and discussing the outcome, their discussion about the "effect" of the car crumpling empowered them to abandon their initial strategy:

(82) J: Then it all falls back to what we said originally, the crash, the speed of the crash dictates if position and dummy time, you know, the crumpling of the car would have an effect.

(83) L: Yeah. I don't understand why we have to do different tests for each three different sections [investigation questions]. You know? You click on them and be like whatever trials for this, kind of. 'Cause it looks like we're kind of figuring it out as we're looking at this.

Here, the *compare* questions promoted a discussion about the relative "effects" of the variables on the outcomes that the *isolate* questions, by their very nature, were unlikely to promote. The discussion about effects appears at least in part to have led Linda and Joann to abandon their initial approach. Unlike Brett and Eric, they had prioritized "figuring it out" over conducting "whatever trials" that were expected of them for each investigation question. At this point, they took a different approach toward making sense of the *Airbags* situation.

### Trials 5 - 7: Exploring extreme values

In trials 5, 6, and 7, Joann and Linda explored extreme values of the velocity and time variables. In trial 5 they simulated a "high impact crash", setting the velocity to the "fastest possible." In trial 6, they "try it with dummy time if we put it at, like, zero." In trial 7, they tested the maximum crumple zone ("All the way?"). This sequence culminated in an important observation Joann made as they examined trials 6 and 7:

(137) J: Well, I guess, dummy time can also have an effect with position, cause like I'm saying, if you have no dummy time, then how close you are to the steering wheel matters a lot.

Again, only the *compare* questions would prompt Joann make this insight about how much the proximity to the steering wheel "matters". In this case, the extreme value approach illustrated to Joann that the effect of time on the outcome may depend on the value of position. Though this initial understanding of the interaction between position and time was at this point incomplete, it became increasingly sophisticated through continued experimentation and discussion, eventually leading to a highly nuanced understanding of the threshold values.

### Trials 8 - 11: Incremental variation

Beginning with trial 8, Joann and Linda changed approaches once more to examine the effects of minute changes to the variables. From trial 8 to trial 9, they decreased the time by the smallest possible increment. After not finding anything conclusive, they incrementally adjusted the velocity in trial 10. Comparing trials 9 and 10 revealed to them nearly identical graphs and led to the following exchange:

(175) L: …So then that one doesn't make much of a difference.
(176) J: No. OK.
(177) L: So the speed didn't make a difference in that—
(178) J: Oh. Doesn't that counteract what we first said? ...I mean, contradict?
…
(190) J: I'm thinking that, he doesn't start moving until, I don't know, I can't put it in words. Even though he's moving faster, by even a little bit, it doesn't really have much of an effect because then, the amount of time….

Observing the similarity between the trial 9 and trial 10 graphs made them realize for the first time that their initial hypothesis (that speed would have the largest effect) was incorrect. Joann was almost able to explain why, but could not quite articulate the precise reason for the insensitivity of the outcome to the velocity. However, using trial 11 to retest the situation once more with the highest allowable velocity value crystallized their understanding of the time threshold:

(207) L: So it's the same, the speed doesn't change anything, it's the same graph we just had.
(208) J: Oh, in that case, cause he won't start moving until—he won't start moving *period* [verbal emphasis] until—this has already been inflated no matter how fast he's going. Yeah. Because he won't start moving until .06 seconds has gone by.
(209) L: Right.
(210) J: So he could be moving at 100,000 miles an hour and he won't hit it until the airbag's already inflated, according to how we set this up.

One revealing aspect of Joann's and Linda's last four trials is that they do indeed constitute use of CVS. However, they conducted these controlled trials in a different way (using incremental variation) and for a much different purpose (to compare the magnitude of effects) than for their earlier controlled trials, or for Brett's and Eric's controlled trials. Not surprisingly, they were also much more informative than the other trials. The design of these trials made use of their previous insights, and their use of CVS emerged spontaneously as the best way to further their understanding of the situation, rather than as a strategy they learned in science class.

After conducting trial 11, Joann's and Linda's understanding of the situation was so complete, they did not require any additional trials to succinctly characterize the relative effects of position, velocity and time on the outcome, on the basis of threshold values:

(227) J: …if we move it closer to the steering wheel, then dummy velocity and dummy time wouldn't matter, because as the airbag starts inflating, he'd be in the way…
…
(238) J: And, he waits that certain amount of time for the airbag to inflate, then velocity doesn't matter.

## Discussion

*Airbags* benefited learners across the distribution of prior knowledge, albeit in different ways. Learners with initially poor understanding of motion graphs made large gains in their abilities to interpret and construct motion graphs, and they were able to generalize their knowledge from *Airbags* to other motion contexts. However, these students' experimentation choices and responses to embedded prompts show that many of them struggled to investigate the questions systematically and reach meaningful insights. On the other hand, learners with initially strong understanding of motion graphs had less room for improvement in that area, but conducted experiments that led to more valid inferences. In particular, the *compare* questions helped the most sophisticated learners examine key distinctions between the variables, such as the magnitudes of the effects of each variable on the outcome.

The stark differences between the two cases illustrate how the *isolate* and *compare* questions led students to reason about the *Airbags* situation. For Brett and Eric, the *isolate* questions appeared to provoke a "schoolish" interpretation of the task. They viewed the task as a simple covariation problem (a common task in school science), and as a result they limited themselves to a predetermined pattern of using CVS rather than spontaneously employing diverse strategies to investigate new questions. They prioritized validly implementing CVS over gaining insight. They sequestered their understanding of the *Airbags* situation from their investigation strategy and, more generally, from their conceptions of the *Airbags* task and the real life practice of science. Though their strategy was "valid" (as judged by criteria often imposed by classroom science), it was not especially informative. Their analysis did not go beyond a superficial characterization of the variables.

In contrast, the *compare* questions prompted Joann and Linda to incorporate a wider range of strategies to elucidate variation patterns. They conducted trials intending to understand the relationships between variables and outcomes and the mechanisms that governed these relationships. Though their initial efforts to use CVS did not yield useful ideas, in the end they spontaneously used CVS to achieve a nuanced understanding of *Airbags* by building on ideas they refined using other strategies. In this way, Joann's and Linda's use of CVS during their final four trials evolved naturalistically as a way to investigate questions that arose from their previous trials. The *compare* questions encouraged Joann and Linda to deeply consider the nature of the variables and to make important and meaningful distinctions between them.

The complexity of Joann's and Linda's analysis illustrates why we observed differences between the conditions only for high prior knowledge students. Joann and Linda needed highly sophisticated knowledge of experimentation strategies and graph interpretation in order to reach their advanced level of understanding. Less sophisticated students likely lacked sufficient knowledge either to distinguish the *compare* questions from the *isolate* questions or to adequately investigate either type of question. Future research will examine how to refine the guidance to help more typical students conduct informative experiments.

## Conclusions and Implications

This study investigated the effect of encouraging students to compare rather than isolate variables. In *Airbags,* students who were guided to *isolate* variables often missed important insights by limiting their analysis to one variable at a time. Prompting students to *compare* variables led to a more nuanced understanding of the *Airbags* situation particularly for high prior knowledge students, achieved at least partly by using diverse experimentation strategies.

Our findings point to the value of providing opportunities for multiple approaches to experimentation rather than guiding students only to isolate variables. The findings suggest that instructional designers should balance guidance designed to promote CVS with opportunities to explore the nature and meaning of the variables. Designers of instruction should select problems where subtle distinctions such as thresholds are necessary for complete understanding. Many everyday problems such as decisions about drug dosage require an understanding of thresholds. Our study demonstrates the value of connecting experimentation to real-life contexts such as airbags, where students can appreciate connections between science instruction and everyday life.

This study also illustrates the value of providing students with graphical representations to help them make valid inferences from their experiments. Our detailed analysis of the ways students reasoned with the *Airbags* visualization revealed that the graphs were essential for comparing multiple trials, revisiting previous ideas, and interpreting collision events. How graphing tools should be designed to support scientific reasoning merits further study.

## References


Buckley, B., Gobert, J. & Horwitz, P. (2006). Using Log Files To Track Students' Model-based Inquiry. In S. A., Barab, K. E Hay & D. T. Hickey (Eds.) *Making a difference: Proceedings of the Seventh International Conference of the Learning Sciences* (pp. 57–63). Mahwah, NJ: Erlbaum.

Clement, J. (1983). Using Bridging Analogies and Anchoring Intuitions to Deal with Students' Preconceptions in Physics. *Journal of Research in Science Teaching, 30*, 1241–1257.

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Kanari, Z. & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*, 748–769.

Klahr, D., Fay, A., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*, 111–146.

Koslowski, B. (1996). Theory and evidence: The development of scientific reasoning. Cambridge: MIT Press.

Kuhn, D. & Phelps, E. (1982). The development of problem-solving strategies. In H. Reese (Ed.), *Advances in child development and behavior* (vol. 17, pp. 1-44). New York: Academic Press.

Linn, M. C. (2005). WISE Design for Lifelong Learning—Pivotal Cases. In P. Gärdenfors & P. Johannsson (Eds.), *Cognition, Education and Communication Technology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C., Davis, E. A., & Bell, P. (Eds.). (2004). *Internet Environments for Science Education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C., Lee, H.-S., Tinker R., Husic, F., & Chiu J. L. (2006). Teaching and assessing knowledge integration in science. *Science, 313*, 1049–1050.

McElhaney, K.W. & Linn, M.C. (2008). Impacts of students' experimentation using a dynamic visualization on their understanding of motion. *International Perspectives in the Learning Sciences: Cre8ting a Learning World. Proceedings of the 8th International Conference of the Learning Sciences.* International Society of the Learning Sciences Inc., Utrecht, The Netherlands.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology. 32*(1), 102–119.

Schauble, L., Klopfer, L.E., & Raghavan, K. (1991). Students' Transition from an Engineering Model to a Science Model of Experimentation. *Journal of Research in Science Teaching, 28*(9), 859–882.

Schwartz, D. L. & Bransford, J. D. (1998). A Time for Telling. *Cognition and Instruction, 16*(4), 475–522.

Siegler, R.S. & Liebert, R.M. (1975). Acquisition of formal experiment. *Developmental Psychology*, *11*, 401–412.

Tschirgi, J. (1980). Sensible reasoning: a hypothesis about hypotheses. *Child Development*, 51, 1–10.


## Acknowledgments

We thank Carol Lee, Brian Reiser, and Bruce Sherin for helpful discussions and three anonymous reviewers for suggestions. We acknowledge the support of National Science Foundation Grant No. ESI-0334199.

# Fading Instructional Scripts: Preventing Relapses into Novice Strategies by Distributed Monitoring

Christof Wecker, Frank Fischer, University of Munich, Leopoldstr. 13, D-80802 Munich
christof.wecker@psy.lmu.de, frank.fischer@psy.lmu.de

**Abstract:** During the fading of instructional scripts learners might relapse into their initial novice strategies after script prompts are withdrawn. One possibility to overcome this problem could be a learning partner providing distributed monitoring of the performance of the strategy suggested by the script. In a 2-factorial experiment with the factors fading and distributed monitoring that involved 126 students of educational science we investigated whether during the fading of an instructional script there is a moderating effect of distributed monitoring on the performance of the strategy and thereby on the acquisition of the cognitive skill of argumentation. Strategy knowledge was fostered best by the combination of fading and distributed monitoring. Distributed monitoring also kept the performance of some aspects of the strategy at higher levels during fading, which was positively related to the acquisition of strategy knowledge.

## Theoretical background

### Argumentation as a cognitive skill

In the literature on cognitive skill acquisition, a skill is typically regarded as a system of procedural knowledge that can be described by a set of production rules (e. g. Newell, 1990; Anderson & Lebiere, 1998). Each single procedural knowledge component described by a production rule can fulfil at least one of two functions: It can regulate the execution of the skill by setting subgoals, or it can directly contribute to performance by helping accomplish these subgoals (cf. Anderson, 1987). The main mechanism for acquiring a unit of procedural knowledge according to these theories is applying it (Anderson & Lebiere, 1998).

Argumentation competence can be regarded as a cognitive skill. On the basis of a cognitive task analysis, the ability to arrive at a counterargument against the relevance of someone else's argument for a specific claim (as one aspect of argumentation competence) can be described as a series of subgoals that can be accomplished by procedural knowledge that implements argumentation schemes (Wecker, 2008): (1) the identification of a claim in someone else's utterance, (2) the identification of an argument put forward to support the claim, (3) the identification of the type of the claim, (4) the identification of the type of the argument, (5) a check of the fulfilment of the conditions of relevance of the argument with respect to the claim, and (6) the formulation of a counterargument on the basis of the results of the analysis conducted in steps 1 to 5. Accordingly, the specific aspect of argumentation competence mentioned above involves strategy knowledge about this series of steps.

Learners may be supported by means of an instructional script to apply such a strategy while engaging in collaborative learning (e. g. Stegmann et al., 2007). As long as they do not master the strategy, they use general-purpose productions to interpret these instructions (Taatgen, Lebiere & Anderson, 2006) and set the corresponding subgoals in accordance with them. In this phase, learners do not yet apply the strategy knowledge themselves since control of performance is exerted completely by the script that guides them. Consequently, this kind of perfomance contributes little to the acquisition of strategy knowledge according to the assumptions of skill acquisition theories mentioned above. In order to strengthen strategy knowledge, learners need to use it to self-regulate their performance. Fading can provide them with opportunities to do so.

### The fading of instructional scripts

Very diverse kinds of instructional support can be faded, ranging from stimuli and prompts (Riley, 1995) to steps in worked-out examples (Renkl & Atkinson, 2003). In most of these cases, fading proved effective for learning (e. g. Schunk & Rice, 1993; Renkl, Atkinson & Große, 2004). However, the kinds of instructional support faded in these studies are quite unlike instructional scripts. These are more akin to scaffolds; collaboration scripts can actually be regarded as a kind of socio-cognitive scaffolding (Carmien, Fischer, Fischer & Kollar, 2007).

Fading has always been regarded as an integral part of scaffolding (Wood, Bruner & Ross, 1976; Pea, 2004; Puntambekar & Hübscher, 2005). So far, only a couple of studies on the effects of the fading of scaffolds have been conducted, with mixed results. Leutner (2000) conducted two experiments on the effects of fading on the acquisition of software skills. One provided evidence for beneficial effects of fading, the other indicated decreased performance in the process of fading. McNeill, Lizotte, Krajicik and Marx (2006) demonstrated a marginally significant positive effect of fading on knowledge about the principles of scientific explanations. In a study by Lee and Songer (2004) the fading of scaffolds had no effect on the quality of explanations in the posttest, but the quality of explanations during learning decreased over time while scaffolds were faded.

From the perspective of cognitive skill acquisition these results do not come as a surprise: As long as there is unfaded support, learners do not practice the application of strategy knowledge to self-regulate their performance. As soon as support is faded, they are immediately required to jump in and exert self-regulation of their performance, which they had no opportunity to practice before. A way out of this paradox may be available in collaborative situations.

### The role of distributed monitoring

When support previously available from an instructional script is gradually withdrawn, successful performance requires learners to take over control of their activities. A full cycle of control involves planning, monitoring and adapting one's steps as part of a strategy. This task may overwhelm learners if they are supposed to take responsibility for all of its parts at once. In such a situation, the idea of distributed metacognition (King, 1998) proves useful: Specific components of control may be distributed among collaborating learners. For example, a learner can be freed from the task of monitoring his or her own application of a strategy and receive feedback on the performance from a learning partner. As soon as he or she wanders off track, such feedback can be used in subsequent cycles to plan the single steps in accordance with the strategy to be internalized.

### Research questions

Accordingly, this study focused on the following research questions:
(1) What are the effects of fading and distributed monitoring on the acquisition of strategy knowledge?
(2) What are the effects of fading and distributed monitoring on the performance of the strategy suggested by the script in the course of time?
(3) What is the relation between the performance of the strategy and the acquisition of strategy knowledge?
It was assumed that fading will play out its full potential to foster the acquisition of strategy knowledge only when combined with distributed monitoring. The hypothesized mechanism behind this effect is self-directed (i. e. unguided) performance of the strategy suggested by the script, which may be kept up by distributed monitoring.

## Methods

### Participants

The participants of the study were 126 students in courses in educational science and teacher preparation who attended a lecture with the title "Introduction to Educational Psychology". On average, they were 23.3 years old (SD = 3.9). Of them, 73.8 % were female and 26.2 % were male. They were randomly paired to dyads who discussed on separated online discussion boards during a collaborative learning phase.

### Design

A 2x2 design with the factors fading and distributed monitoring was implemented (see table 1).

Table 1: Design of the study.

| **Distributed monitoring** | **Fading** | |
|---|---|---|
| | *Unfaded script* | *Faded script* |
| *Without distributed monitoring* | 15 dyads/30 persons | 18 dyads/36 persons |
| *With distributed monitoring* | 13 dyads/26 persons | 17 dyads/34 persons |

### Learning task, material and environment

The two learners in each group dealt with cases on the application of Weiner's attribution theory in a text-based online discussion board. They were told that they were discussing analyses of these cases in groups of four and that two of the other learners had the task to write these analyses. The learners themselves were asked to write critical replies to each of theses analyses and could discuss any questions that came up during this task on the discussion board. In fact, six case analyses with at least two questionable claims were posted to the board under the names of two simulated group members at fixed points of time. As a preparation for this collaborative learning session, the learners read a three-page text on Weiner's attribution theory and a four-page text on how to generate counterarguments for the critical replies.

A script supported the learners in all four conditions in the process of formulating counterarguments against the prepared case analyses by providing instructions on how to analyze the argumentation in the case analyses to discover problematic assumptions. The script contained three kinds of information: sequence information, argument schemata and application support (see figure 1). *Sequence information* describes the process of analyzing the argumentation in the prepared case analysis and the construction of a critical reply to it. The steps of the strategy to arrive at a counterargument were mirrored in the interface as prompts that changed according to the state of the editing, and specified the next step. *Argument schemata* contained information on what types of argument are appropriate to support the identified type of claim and what conditions of relevance need to be fulfilled for an identified pair of argument and claim, which was crucial for the assessment of the argumentation in the case analysis. They were implemented by means of selection fields for the type of the claim and the argument as well as by a prompt for the assessment of the conditions of relevance. *Application support* was provided by means of explanatory sentences for the terms used in the prompts and selection fields of the script as well as examples for the respective types of propositions. They were displayed in the interface directly next to the respective control elements.

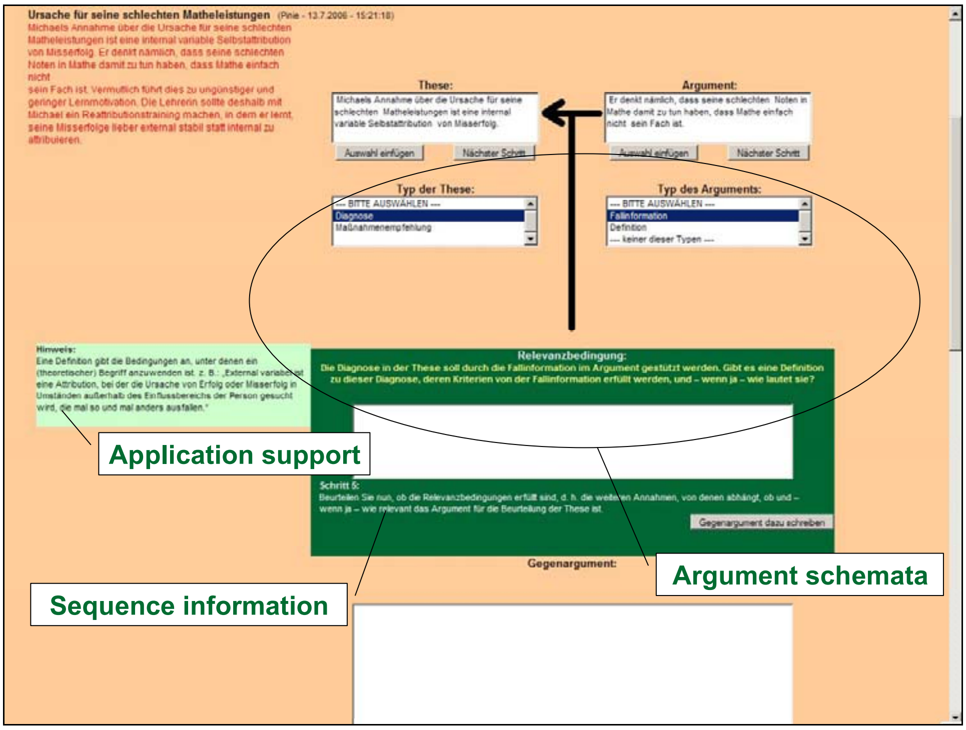



Figure 1. Implementation of the script for the formulation of counterarguments.

## Independent variables

### Fading

In the conditions *with faded script* components of the interface were removed based on the number of critical replies posted. The following fading regime was used: The *application support* disappeared completely after two critical replies had been written. *Sequence information* prompts were faded by replacing two randomly chosen prompts per round by the request “Please perform this step on your own.”, starting with the composition of the third critical reply. This means that after five critical replies only this unspecific request was shown before each step. With respect to the *argument schemata* “later” elements had to be faded first because for the branching of the support for different argument schemata unequivocal input was required: First the specific question concerning the fulfilment of the conditions of relevance (step 5) was replaced by an unspecific one. From the forth critical reply on, the dropdown field for the type of the argument did no longer contain any options, but the learners had to fill in the type of the argument themselves. From the fifth critical reply on, the dropdown field for the type of the claim did no longer contain any options, but the learners had to fill in the type of the claim themselves. After five critical replies had been posted, the interface did not change as a function of the number of critical replies any more, no matter how many messages were written. After 70 minutes, finally the students were provided only with a simple text box for the formulation of their critical replies to the case analyses as customary in asynchronous discussion boards.

In the conditions *with unfaded script* the interface for the composition of critical replies remained unchanged throughout the learning phase.

### Distributed monitoring

In the conditions *with distributed monitoring*, one of the learnings had the task to provide the other one with feedback for each of his or her critical replies, based on which the other learner was asked to revise his or her critical reply. During the formulation of this feedback, the learning partner was supported by the interface: By simply clicking on check boxes, a message about the completeness of the six steps of the strategy for the construction of a counterargument, on the appropriateness of the identification of the types of claim and argument, and on the correctness of the answer to the question concerning the conditions of relevance could be composed. Furthermore there was the opportunity to add text remarks. In the condition with faded script and distributed monitoring, the distributed monitoring was continued after the fading had started.

In the conditions *without distributed monitoring*, the learning partners were neither asked nor supported to provide feedback on the critical replies.

## Dependent variables

### Strategy knowledge

Strategy knowledge was measured by means of a task with open answering format. The learners were asked to describe their strategy for checking the relevance of an argument for a claim and formulating a counterargument against it. The learners' unsegmented answers were coded for the occurence (0 – absent/1 – present) of each of the six steps of the strategy implemented in the script. One further coding item captured the correctness of the sequence. The agreement of two coders for these coding variables ranged from 76 % to 90 % (median: 86 %); Cohen's $\kappa$ ranged from .46 and .70 (median: .51). Therefore, the objectivity of the codings can be regarded as sufficient (cf. Orwin, 1994, p. 152). The seven coding items were added up to form the scale for strategy knowledge with a possible range from 0 to 7. The seven items were internally consistent (Cronbach's $\alpha$ = .93).

### Performance of the strategy

Performance of the strategy was measured based on five single variables indicating for each critical reply for each of five of the six steps of the strategy whether this step had been performed (1) or not (0). The last step was omitted since it coincided with composing the message, which is performed in any reply irrespective of the strategy applied. For the first four steps, the corresponding adherence variables were taken directly from logfiles. The adherence variable for the fifth step of the strategy was coded (percentage agreement: 99 %, Cohen's $\kappa$ = .98).

The overall scale for the performance of the strategy was formed in two steps: First, for each of the five steps the proportion of all messages in which the step had been performed was calculated (also separately for all messages before and after the fading of support). Finally, the indicator for overall performance of the strategy was computed by adding the five variables for the performance of the single steps and could range from 0 to 5. Its reliability was rather high (Cronbach's $\alpha$ = .86).

## Procedure

Data were collected in a series of sessions of three hours of length with up to 20 students each. These were distributed over two rooms; learning partners who collaborated online sat in different rooms. After a short introduction into the purpose and procedure of the study, the participants filled in an online questionnaire for control variables and read printed texts on attribution theory and on how to construct counterarguments. They could keep them until the end of the learning phase. The collaborative learning phase started with a demo video on how to use the learning environment. After a short break, an 80-minute collaboration phase in the different experimental conditions followed. Finally, the learners completed online post-tests.

## Statistical analysis

Data were analyzed with individual students as the units of analysis. From each dyad, one member was selected who had not provided distributed monitoring. Furthermore, a hierarchical-linear analysis of the performance of the strategy over the course of time was conducted with single messages (the critical replies) as the units of analysis. The significance level was set to 5 % for all analyses.

# Results

## Effects of fading and distributed monitoring on the acquisition of strategy knowledge

The descriptive results for the acquisition of strategy knowledge in the four conditions are displayed in table 2. An analysis of variance with strategy knowledge as the dependent variable and fading and distributed monitoring as independent factors shows that strategy knowledge was significantly higher in the faded script conditions than in the unfaded script conditions, which corresponded to a medium size effect of fading, $F(1; 62) = 6.32$; $p < .05$; partial $\eta^2 = .09$. Likewise, a significant medium to large size effect of distributed monitoring on strategy knowledge in favour of the conditions with distributed monitoring compared to the conditions without distributed monitoring was detected, $F(1; 62) = 8.20$; $p < .01$; partial $\eta^2 = .12$. The small to medium size interaction effect of these two independent variables was marginally significant, $F(1; 62) = 3.21$; $p < .10$; partial $\eta^2 = .05$. However, the main effects are largely due to the superiority of the condition with the faded script along with distributed monitoring. Students in this condition significantly outperformed the students in the three other conditions on the strategy knowledge test.

Table 2: Means and standard deviations of strategy knowledge in the experimental conditions.

| | | | *Strategy knowledge* | |
|---|---|---|---|---|
| *Distributed monitoring* | *Fading* | *N* | *M* | *SD* |
| Without distributed monitoring | Unfaded script | 15 | 0.79 | 1.81 |
| | Faded script | 18 | 1.47 | 2.13 |
| With distributed monitoring | Unfaded script | 13 | 1.46 | 2.30 |
| | Faded script | 17 | 4.24 | 2.88 |

## Effects of fading and distributed monitoring on the performance of the strategy suggested by the script

Research question 2 asked whether fading and distributed monitoring affect the performance of the strategy suggested by the script in the course of time . The temporal development of the performance of the strategy was analyzed by means of a hierarchical-linear analysis. The level-1 growth model predicted the performance of the strategy during the formulation of message t from person i on the basis of the temporal position of message t:

$$y_{ti} = \pi_{0i} + \pi_{1i} \cdot (\text{number of message}_t) + e_{ti}$$

Intercepts (base levels) and slopes (growth rates) varied substantially between students, intercepts: $\chi^2(62) = 229.73$; $p < .001$; $\rho = .34$; slopes: $\chi^2(62) = 120.79$; $p < .001$; $\rho = .02$. As can be seen from the intercept for number of message presented in table 3, however, on average there was neither an increase nor a decrease in the performance of the strategy suggested by the script over time.

The explanatory level-2 model used the experimentally manipulated variables fading and distributed monitoring to predict both the base level (intercepts $\pi_{0i}$) and the growth rates (slopes $\pi_{1i}$) of the performance of the strategy for person i:

$$\pi_{0i} = \beta_{00} + \beta_{01} \cdot (\text{fading}_i) + \beta_{02} \cdot (\text{distributed monitoring}_i) + r_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11} \cdot (\text{fading}_i) + \beta_{12} \cdot (\text{distributed monitoring}_i) + r_{1i}$$

As can be seen from table 3, the base level did not vary significantly as a function of either fading or distributed monitoring. The average individual growth rate in the condition with the unfaded script without distributed monitoring (basal growth rate), however, was significantly below zero, indicating a decrease in performance of the strategy in this condition. While fading did not significantly affect this growth rate, distributed monitoring significantly raised this negative growth, yielding an approximately constant level of performance of the strategy in the corresponding conditions ($-0.08 + 0.11 > 0$). The proportion of variance in slopes accounted for by the explanatory model was 6 %, which was still significant, $\chi^2(62) = 117.59$; $p < .001$, indicating that there may be further factors that substantially contribute to it.

Table 3: Hierarchical-linear analysis of the development of the performance of the strategy triggered by the script as a function of fading and distributed monitoring.

| *Growth model: Prediction of the performance of the strategy $y_{ti}$* | $\pi_1$ | *t* | *p* |
|---|---|---|---|
| Number of message | -0.03 | -1.18 | .24 |
| *Explanatory model:* | | | |
| *Prediction of the base level $\pi_{0i}$ in the growth model* | $\beta_{0q}$ | *t* | *p* |
| Fading | 0.21 | 1.39 | .17 |
| Distributed monitoring | 0.09 | 0.57 | .57 |
| *Prediction of the growth rate $\pi_{1i}$ in the growth model* | $\beta_{1q}$ | *t* | *p* |
| Basal growth rate (without fading and distributed monitoring, *intercept* $\beta_{10}$) | -0.08 | -2.19 | .03 |
| Fading | 0.01 | 0.24 | .81 |
| Distributed monitoring | 0.11 | 2.45 | .02 |

The role of distributed monitoring for the performance of the strategy suggested by the script was analyzed separately in the conditions with the faded script by means of an analysis of variance for repeated measures with performance of the strategy as the dependent variable, distributed monitoring as a between-subjects factor and fading as a within-subjects factor with the two values before fading and after fading, i. e. the performance of the single steps of the strategy was averaged over all messages from each participant written before support for the particular step was faded and over all messages written after support for the particular step was faded, and aggregated over the five steps of the strategy (see figure 2).

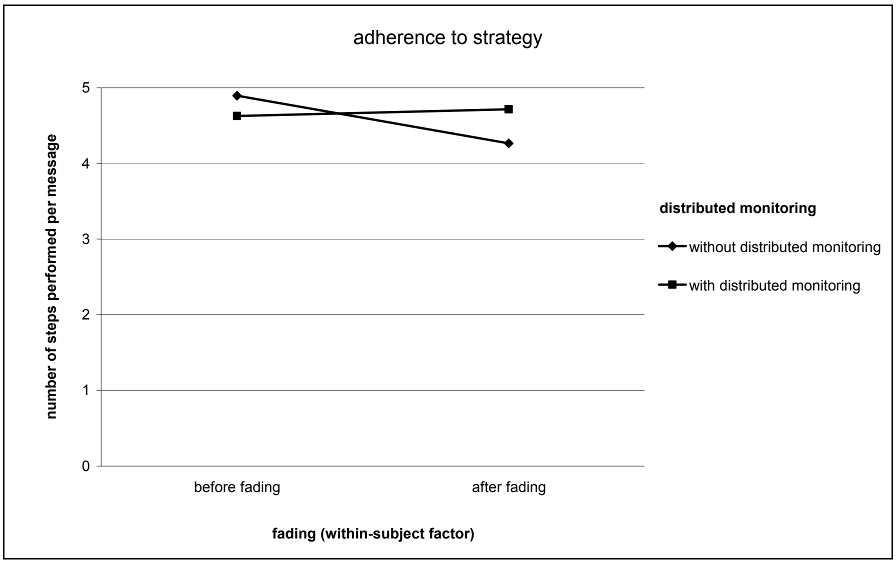



Figure 2. Average proportions of messages in which the strategy as a whole was performed before and after fading (separated for the conditions with and without "distributed monitoring").

A decrease in the performance of the strategy was found that reached the boundary of significance, $F(1;\ 31) = 4.16$; $p = .05$; partial $\eta^2 = .12$. However, this decrease occurred only in the condition without distributed monitoring, as indicated by a significant interaction between the within-subjects factor fading and distributed monitoring, $F(1;\ 31) = 7.38$; $p < .05$; partial $\eta^2 = .19$.

Exploratory analyses on the level of individual steps of the strategy revealed that these effects were mainly due to the learners' performance of the last two steps of the strategy: Both for the identification of the type of the argument and the check of the fulfilment of the conditions of relevance there was a decrease after the fading of the corresponding support in the condition without distributed monitoring, identification of the type of the argument: $t(15) = 1.97$; $p < .05$ (one-sided), check of the fulfilment of the conditions of relevance: $t(15) = 2.32$; $p < .05$ (one-sided). For the identification of the type of the argument the interaction effect between the within-subjects factor fading and distributed monitoring was marginally significant, $F(1;\ 31) = 3.06$; $p < .10$; partial $\eta^2 = .09$, while for the check of the fulfilment of the conditions of relevance the corresponding interaction was significant $F(1;\ 31) = 5.27$; $p < .05$; partial $\eta^2 = .15$.

## Relation between the performance of the strategy and the acquisition of strategy knowledge

Between the performance of the strategy and the acquisition of strategy knowledge there was a significant small correlation, $r = .22$; $p < .05$. Because on average there were no differences in the performance of the strategy between the four experimental conditions, this finding does not explain the differences in the acquisition of strategy knowledge between the experimental conditions.

As argued before, the performance of a strategy while learners are guided through its single steps rather than performing them in a self-directed way, should contribute little to the acquisition of strategy knowledge. What is expected to contribute to the acquisition of strategy knowledge is the performance of the strategy in phases in which support has been withdrawn and learners have the opportunity to practice the self-regulation of these steps, which occurred to a larger extent when they received distributed monitoring, as described in the previous section. Accordingly, separate correlations between the performance of the single steps *before and after* the fading of the corresponding support and strategy knowledge about these specific steps were calculated. These are reported for the last two steps because, as described above, for these two there were differences in performance in the course of time between conditions with and without distributed monitoring: Performance of the step of the identification of the type of the argument before the fading of the corresponding support and strategy knowledge about it were not significantly correlated, $r = -.15$; *n. s.*, whereas performance of it after the fading of the corresponding support was marginally correlated with the corresponding knowledge, $r = .30$; $p < .10$. Most importantly, the difference between these two correlations was significant, $z = 1.71$; $p < .05$. Similarly the performance of the step of checking the fulfilment of the conditions of relevance before the fading of the corresponding support and strategy knowledge about this step were uncorrelated, $r = .00$; *n. s.*, while performance of it after the fading of the corresponding support again was marginally correlated with strategy knowledge about it, $r = .33$; $p < .10$. In this case, the difference between the two correlations was marginally significant, $z = 1.59$; $p < .10$.

## Discussion

The main findings of this study indicate that the performance of the strategy implemented in a script may decline over time, especially if components of the script are faded. It could be demonstrated however, that it can be kept on a continuously high level by means of peer support such as distributed monitoring. The performance of the strategy was shown to be related to the acquisition of strategy knowledge, particularly in phases in which components of the scripts have been faded and accordingly the learners have the opportunity to practice the self-regulation of the steps of the strategy. This provides an explanation for the beneficial effect of the combination of a faded script with distributed monitoring on the acquisition of strategy knowledge.

The current study extends our understanding in several ways. With respect to the effects of collaboration scripts, it replicated the finding that scripts are appropriate means to trigger specific learning activities (Weinberger et al., 2005; Kollar et al., 2007; Stegmann et al., 2007), even after the fading of components of the script. It could be shown, however, that the level of performance of the script can be kept at an even higher level by providing distributed monitoring.

With respect to fading, thereby an important condition of its effectiveness was identified: Learners may need further support to take advantage of the opportunity to self-regulate their performance. This finding might contribute to a clarification of the reasons for the varying effects of fading reported in previous studies on fading (cf. Lee & Songer, 2004; McNeill et al., 2006; Leutner, 2000).

This opportunity to practice the self-regulation of the steps of a strategy may not be sufficient if learners fail to take advantage of it. Distributed monitoring can be regarded as one potential way to raise learners' self-regulatory activities in accordance with the strategy implemented in a script even after the fading of the components to support it. Accordingly, distributed monitoring can be one way to make use of collaboration during the acquisition of complex skills in computer-supported collaborative learning. It is also one option for adapting support to the current needs of learners because the learning partner providing distributed monitoring will only jump in with corrective feedback when there is divergence from the strategy to be internalized.

This study went beyond standard practice in research also in a methodological respect by analyzing learning activities diachronically by means of multi-level analyses of the temporal development of specific quality aspects of learning activities instead of aggregating indicators of the quality of learning activities over the whole learning phase (e. g. Weinberger et al., 2005). Thereby it can more closely account for the causal structure of the single learning events influenced *by the independent variables*. This should be pushed even further by relating single learning events to learning outcomes since the way in which learning activities might affect *learning outcomes* was still analyzed on the basis of such aggregated indicators. How this can be done is a topic for further discussion.

It is an important limitation of the present study that argumentation skill was not measured on the basis of perfomance in argumentative situations but by means of a declarative test of strategy knowledge about how to arrive at a counterargument against a position advanced by a learning partner. Furthermore, the study captured only rather short term effects of learning with a faded script and distributed monitoring. As it was conducted under laboratory conditions, the claims put forward in this paper still are in need of validation for more natural learning environments.

Accordingly, future research should focus on effects of fading and distributed monitoring not only on declarative strategy knowledge but also its application in the context of executing argumentation skill. Effects of longer-term interventions on immediate learning outcomes as well as their retention over more extended periods of time should be studied, preferably in field contexts with more authentic kinds of collaboration. Further research should also test the theoretical claims put forward in this study with other domain-general learning outcomes such as online search competence, which is currently being undertaken (Wecker, Kollar, Fischer & Prechtl, 2010). Finally, ways to adapt the fading of a script to a learner's current competence level would be a very promising direction to explore.

Based on the findings from this study it can be recommended that scripts should be faded out to provide learners with the opportunity to practice the self-regulation of skilled performance (as suggested by Pea, 2004; Puntambekar & Hübscher, 2005; and others). This recommendation has to be supplemented by the caveat, however, that it is important to keep learners' performance of the strategy to be acquired at a high level in these self-regulated phases. Collaboration may be exploited to accomplish this goal.

Thus, fading may be a way to move from a high degree of support to self-directed learning with authentic tasks. In this process the acquisition of competence can be considered as an internalization of control that has been exerted socially by peers before.

## References


Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94*(2), 192-210.

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Carmien, S., Fischer, F., Fischer, G. & Kollar, I. (2007). The interplay of internal and external scripts - a distributed cognition perspective. In F. Fischer, H. Mandl, Haake & I. Kollar (Eds.), *Scripting computer supported communication of knowledge: Cognitive, computational and educational perspectives* (pp. 303-326). New York: Springer.

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review, 10*(1), 57-74.

Kollar, I., Fischer, F. & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708-721.

Lee, H. S. & Songer, N. B. (2004). *Expanding an Understanding of Scaffolding Theory Using an Inquiry-Fostering Science Program*. Unpublished manuscript.

Leutner, D. (2000). Double-fading support – a training approach to complex software systems. *Journal of Computer Assisted Learning, 16*, 347-357.

McNeill, K. L., Lizotte, D. J., Krajcik, J. & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences, 15*(2), 153-191.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Orwin, R. G. (1994). Evaluating coding decisions. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 139-162). New York, NY: Russell Sage Foundation.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13*(3), 423-451.

Puntambekar, S. & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist, 40*(1), 1-12.

Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15-22.

Renkl, A., Atkinson, R. K. & Große, C. S. (2004). How fading worked solution steps works – a cognitive load perspective. *Instructional Science, 32*, 59-82.

Riley, G. A. (1995). Guidelines for devising a hierarchy when fading response prompts. *Education and Training in Mental Retardation and Developmental Disabilities, 30*(3), 231-242.

Schunk, D. H. & Rice, J. M. (1993). Strategy fading and progress feedback: Effects on self-efficacy and comprehension among students receiving remedial reading services. *Journal of Special Education, 27*(3), 257-276.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 421-447.

Taatgen, N. A., Lebiere, C. & Anderson, J. R. (2006). Modeling paradigms in ACT-R. In R. Sun (Ed.), *Cognition and multi-agent interaction: From cognitive modeling to social simulation* (pp. 29-52). Cambridge: Cambridge University Press.

Wecker, C. & Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: The role of distributed monitoring. In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *[Proceedings of the] CSCL 2007* (pp. 763-771). International Society of the Learning Sciences.

Wecker, C., Kollar, I., Fischer, F. & Prechtl, H. (2010). *Fostering Online Search Competence and Domain-Specific Knowledge in Inquiry Classrooms: Effects of Continuous and Fading Collaboration Scripts.* Paper presented at the 9th International Conference of the Learning Sciences (ICLS), Chicago, June 29 – July 2 2010.

Weinberger, A., Ertl, B., Fischer, F. & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*(1), 1-30.

Wood, D., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 17*, 89-100.


## Acknowledgments

This study was partially funded by Deutsche Forschungsgemeinschaft (DFG).

# Student Learning Through Journal Writing in a Natural Science Course for Pre-Elementary Education Majors.

Michael T. Dianovsky and Donald J. Wink, University of Illinois at Chicago
845 W. Taylor St., Chicago, IL 60607
Email: dianovsk@uic.edu; dwink@uic.edu

**Abstract:** This paper describes the use of journals in a general education chemistry course for elementary education majors. In the journals, students describe their understanding of a topic, its development, and its connection to their lives. The following types of reflections were identified in student journals: action, prior knowledge, project ideas, text resources, classroom events, and monitoring of knowledge. A multiple linear regression analysis of total course points earned versus student GPA, ACT score, number of reflections made for each reflection type, and total number of reflection made throughout all journals was used to determine the significance of the type of reflections made by students in relationship to their chemistry content understanding. The results indicate that students who reflected more on classroom events and those who monitored their knowledge correlated positively with a higher content knowledge. However, reflecting on text resources had a negative correlation to students' overall chemistry content knowledge.

## Introduction & Theoretical Background

This paper reports an investigation of student journal writing as a regulator of student metacognition in a general education chemistry course. Journal writing provides an opportunity for learners to understand their own learning process, to increase their active involvement in learning, and to gain personal ownership of their learning (Moon, 1999). Writing journals encourages self-thought, which allows for the development of metacognitive practices and the promotion of new learning by evaluating prior experiences. The use of journal writing in the science classrooms allows students to reflect on their actions, prior experiences, text resources, classroom events, and current knowledge, with reflection referring to "those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciations" (Boud, Keogh, & Walker, 1985). Also, reflection in writing has a strong metacognitive component that helps students monitor their learning (Bangert-Drowns, Hurley, & Wilkinson, 2004). Previous research indicates the importance of having students reflect upon their current knowledge and understanding (Brown, Bransford, Ferrara, & Campione, 1983; Puntembekar & Kolodner, 1998; Davis, 2003). Writing assignments incorporating reflection on content knowledge enables students to formulate a better understanding of unprocessed materials and events and to document how their learning occurs, increasing connections with prior knowledge. Also, using productive reflection in the classroom promotes the knowledge integration process by expanding ideas and allowing students to identify the weaknesses in their knowledge. The nature of the assignment and the prompts given to a student are import for this. Davis (2003) found that students who reflected on generic prompts had a more coherent understanding as they worked on complex projects than those students that reflected on direct prompts.

The present study examines a chemistry course for pre-elementary education majors that includes journal writing assignments in which students respond to generic prompts to present their understanding of certain key topics and to reflect upon their own experiences with the topic. Our study focuses on the types of reflection students exhibited in their journals and how these relate to student content knowledge in the course. We also studied the choices students make with generic prompts that are aimed at developing a metacognitive perspective in their writing. The theoretical framework for this study is composed of metacognition and writing-to-learn perspectives. Metacognition refers to the knowledge concerning one's own cognitive processes (Flavell, 1979). Brown (1987) related metacognition to the learners' knowledge, awareness and control of the processes by which they learn. It was determined that metacognitive learners are able to recognize, evaluate, and reconstruct existing ideas (Baker, 1991). These findings of Brown and Baker represent two views of metacognition currently known as the knowledge and regulation of cognition. Learning is said to occur when new data or information presented to a student conflicts with what he or she already knows and give the student a chance to change their mind to think differently (Cracolice, 2005).

Several different kinds of assignments have been used to promote metacognition in chemistry learning (Rickey & Stacey, 2000). This included the deliberate use of a thinking frame (Mattox, Reisner, & Rickey, 2006), student documentation of their thinking through heuristics such as concept maps (Novak & Gowin, 1984), student work in extended problem solving situations (Cooper *et al*, 2008), and working in using extended writing within laboratory work (Greenbowe & Hand, 2005; Wink & Choe, 2008). The theoretical framework of writing to learn is used in this work because it characterizes the components involved when students respond to an opportunity to write about scientific topics. The writing to learn framework has been used in work by Britton (1970) and Emig (1977). In this approach, learning is supported and assessed by students' writing about their developing ideas. One of the most common forms of writing to learn is done by informal personal journal responses related to educational experiences such as literary readings, scientific experiments, videos, or from discourse with others (McCrindle & Christensen, 1995). The process of writing journals helps with the process of understanding relationships among ideas, knowledge transformation and construction of new meaning (Wallace, 2007; Klein, 1999). A major goal of the course was for students to be able to construct ideas and generate connections between concepts to make sense of the ideas rather than simply memorizing isolated facts. Another goal was to improve students' scientific literacy by having them write about the learning of science and to consider how they process information. The writing to learn strategy will help students achieve these goals. It has been shown that students develop conceptual growth through the process of writing, including parallel metacognitive growth in the learner (Mason & Buscolo, 2000).

In the course, students experienced writing-to-learn in the context of their journal work. Journals are well known in the K-12 literature, including the work of Davis (2003) that also contributes to the methodology of our work. She showed that prompts for student reflection promoted knowledge integration which occurs as learners expand their repertoire of ideas, identify weaknesses in their knowledge, and combine knowledge integration processes together. Journal writing is also used in the college science classroom, for example in a study of reflective journal writing within a first year university biology class (McCrindle & Christense, 1995). Students were randomly assigned to either write reflectively in their "Learning Journal," or produce a scientific report every week during the last five weeks of the course. The students in the journal writing group were given instructions and prompts as to the form of their journal writing. Results indicated that those students who did write reflectively on the content of their learning and the process of their leaning reported a more sophisticated conception of learning and had a direct impact on their metacognitive awareness. Similar research on lab report writing produced similar results (Keys *et al*, 1999).

## Method

### Participants

The participants consisted of 76 students from two offerings of the course (Fall 2007 and Spring 2009), called *The Chemical World*, who gave consent for their work to be included in a research program. Different instructors taught each course. Virtually all of these participants were enrolled in the pre-service elementary education program of the university and they took the course as part of a natural science general education requirement (Varelas *et al,* 2008), which included this course. Out of the seventy-six participants seventy-four were females and two were males, reflecting the existing gender distribution in the course. Students attend a one-hour lecture twice a week held by the instructor and a three-hour laboratory section led by a graduate student. Students are assessed by a student introductory essay, six student journals, student portfolios, unit and final exams, lab reports, assigned topical projects, and a big theme project (Wink *et al*, 2009). The present study focuses on the journal assignments, which are the major place for students to describe developments in their understandings of chemical content.

### Task and Procedure

The students submitted a total of six journals, approximately biweekly, throughout the course. The journal topics that were assigned consisted of material the students covered in lecture or lab in their previous weeks of class. The approximate length of each journal was 700-900 words. The journal entries on various topics of chemistry provided an opportunity for students to reflect on their current understanding and how they came to this understanding. They were also a place where students communicated with and received feedback from the instructor. The prompts for the journals were consistent throughout the semester, with only a single assignment sheet given out at the beginning of the course. Hence, the prompts fit the category of "generic," (Davis 2003), permitting students to work in different ways within a simple scaffold. Students were instructed to include five components in their journals and were given a rubric documenting the layout and value of each component. The first three components all related to a content topic in the course and are the focus of this research. The first component prompts enabled the students to write about their current understanding of a topic. Here the students were able to express, in writing, what they know about the topic and how they know it. The second component required the students to present specific examples of chemistry and chemical substances in relation to the topic. Following this, the third journal component prompted students to make connections by writing about what the topic might mean to them in their personal history, current experience, or future plans, with frequent reference to their everyday life or career choice.

The fourth and fifth components of the journals covered other aspects of their work and are outside of the scope of this paper. Specifically, the fourth component required that students write about their ongoing work on a "big theme project," developing an idea that would become part of an individual final course project reflecting an aspect of chemistry of interest to them (Wink *et al*, 2009). Finally, students wrote a section of their journal reporting their progress in the course. In this section students expressed the concepts or skills that they find unclear or struggle with, how they were progressing, what was working for them, and what external factors were causing them trouble. They were told that a good journal would consist of approximately 700-900 words total, though many students exceeded that number.

Student journals were first coded for the types of reflection on understanding of the six journal topics. The coding scheme for journals is shown in Table 1. This was created from a coding scheme for student reflection in journals that was derived from Davis' study of student reflections on generic and direct prompts (Davis, 2003). This enabled us to code for the *types* of reflections found within the student journals. Each reflection written by the student was coded with the non-hierarchical indices 1-6 to note the type of reflection. Two coders independently coded students' journals for reflections. There was a 92% interrater-reliability between the two coders.

Table 1: Coding for Student Reflection in Journal Responses

1. Reflection on actions: involves thinking about what ways of behaving are most appropriate, as well as contemplating very general goals.
2. Reflection on prior knowledge: involves thinking specifically about prior experiences in which knowledge was gained.
3. Reflection on project ideas: involves thinking about the conceptual ideas presented in the project itself (Mineral, Nutrition, and Big Theme)
4. Reflection on text resources: involves thinking specifically about information presented in the textbook, lab instructions, or handouts.
5. Reflection on classroom events: includes thinking rendered from individual experiences from lectures, labs, and discussions from peers, teaching assistants, and the instructor.
6. Reflection on knowledge: indicates a focus on monitoring or improving one's understanding

Lastly, during the second course offering, students completed the Metacognitive Activities Inventory (MCA-I) developed by Cooper, Sandi-Urena and Stevens (2008) at the end of the semester. This survey represents a way to see if students are metacognitive during problem solving tasks in chemistry. The results of this survey and the reflection made by the students will be compared to see if those students who are actively metacognitive during problem solving are also metacognitive during writing.

## Data Analysis and Findings

### Journals

Throughout the semester students wrote a total of six journals. Six different reflections were apparent in the journals after reading through them, listed in Table 1. All journals were coded for these reflection types. Examples of student reflection responses for the six coded categories are shown in Table 2.

Table 2: Examples of student reflection code

| Reflection Type | Example from Student Work |
|---|---|
| Action | I find if I take the extra time to figure out how to do each problem and read the chapters, I could be well off and prepared. The only problem is managing enough time to set aside for chemistry. |
| Prior Knowledge | Over the last weeks we have learned unit conversions in the metric system in class. I grew up in Germany and used the metric system all my life and am used to the system. It is easier to use and convert in the metric system because it is all powers of ten. I am confident in my understanding of the common prefixes used. |
| Project Ideas | I think that in order to do my individual project it is necessary to know how to work with unit conversions because this will help me understand better how much of each chemical is put in food to make it grow faster and better. |
| Text Resources | I came to understand the properties about polarity and functional groups of organic compounds through reading chapter 9 in Chemistry for Changing Times. Reading the book gave me a basic idea about what functional groups are and what it means to have polarity. |
| Classroom Events | The labs have helped me understand better how scientists work because I have to follow the same procedures/steps they use in order to complete the labs. In order to do the labs I have to think like a scientist and this is very interesting and fun. |
| Knowledge | I feel very confident about my progress in class so far because I am been able to understand how the cycles work and their importance to our survival. I am able to analyze the cycles and point out where each of the four spheres takes place and why. |

## Regression Analysis

To understand how the journals might impact learning, we did a multiple linear regression analysis of the sum of all points earned in the course as the dependent variable against student: GPA prior to course enrollment, ACT composite score, the number of reflections made for each of the six codes in the six journal entries, and the total number of reflections made in each of the six journal entries as the independent variables. This analysis sought to determine which variables correlate significantly with student content knowledge of chemistry. The analysis was done for each of the two classes and an aggregate of both classes. Before the two class sets of data were combined and analyzed, a comparison was done between them to see if there were any significant differences between the students in the two courses. An independent t-test was conducted on prior GPA, ACT composite score, and course grades given to the students. There was no significant different between the two classes in prior GPAs ($t(74) = 0.140$, $p > 0.05$), ACT ($t(74) = -1.210$, $p > 0.05$) and course grades ($t(74) = -0.743$, $p > 0.05$). Each class was given the same set of assignments and covered the same chemistry material. Student course grades derived from the total number of points the student earned during the semester. Both of the two classes had a total of 1075 points that a student could earn. The students' GPA that was used for this analysis was their GPA prior to the semester in which they took the course.

The linear regression model for all three data sets (Fall 2007, Spring 2009, and Aggregated Data) shows a relation between student course grade and four predictor variables, shown in Tables 3-5. This signifies that between 58% and 63% of the variance in student course grade is explained by student prior GPA and three types of reflections made by students in journal writing. The first predictor variable is the students' prior GPA. A high prior GPA is a good prediction of doing well in the class. Student reflection on classroom events (Code 5) and on their knowledge (Code 6) correlated positively and significantly with higher performance in the course. Student reflection on textbook, handouts, and worksheets (Code 4) correlated negatively and significantly with higher course performance.

Table 3: Fall 2007 regression analysis for prediction of course grade based on student GPA, ACT, reflection type, and total number of reflections during Fall 2007

| Significant Predictor Variables | β |
|---|---|
| Prior GPA | .731 |
| Code 4 | -.144 |
| Code 5 | .052 |
| Code 6 | .086 |

*Note:* Regression equation: Grade = .409 + .731(GPA) - .144(Code 4) + .053(Code 5) + .086(Code 6), ($R = 0.794$, $R^2 = 0.630$)

Table 4: Spring 2009 regression analysis for prediction of course grade based on student GPA, ACT, reflection type, and total number of reflections during Spring 2009

| Significant Predictor Variables | β |
|---|---|
| Prior GPA | .551 |
| Code 4 | -.020 |
| Code 5 | .038 |
| Code 6 | .034 |

*Note*: Regression equation: Grade = .593 + .551(GPA) - .020(Code 4) + .038(Code 5) + .034(Code 6) ($R = 0.764$, $R^2 = 0.584$)

Table 5: Aggregated data regression analysis for prediction of course grade based on student GPA, ACT, reflection type, and total number of reflections

| Significant Predictor Variables | β |
|---|---|
| Prior GPA | .590 |
| Code 4 | -.031 |
| Code 5 | .033 |
| Code 6 | .075 |

*Note*: Regression equation: Grade = .315 + .590(GPA) - .031(Code 4) + .033(Code 5) + .075(Code 6) ($R = 0.783$, $R^2 = 0.613$)

## MCA-I Survey

The Spring 2009 class of the *Chemical World* was given the Metacognitive Activities Inventory (MCA-I) survey to see if students who were metacognitive during problem solving were also metacognitive in their writing. This 27 item self-report instrument assesses students' metacognitive skillfulness when solving chemistry problems and may be used as a diagnostic tool in deciding to implement interventions by the instructor (Cooper, Sandi-Urena and Stevens, 2008). Students select their agreement with the items from a 5-point Likert scale (1, strongly disagree to 5, strongly agree). The score was reported as a percentage of the maximum number of points attainable. The higher the percentage a student received relates to a higher metacognitive awareness the student has during problem solving activities. The percentage of the maximum number points for each student was correlated against each of the six reflection types, and the total number of points the student received in the class. The results of these correlations are shown in Table 6.

Table 6: Correlations of MCA-I Survey

| Variables | MCA-I Score |
|---|---|
| Code 1 | -0.041 |
| Code 2 | -0.034 |
| Code 3 | 0.061 |
| Code 4 | -0.436** |
| Code 5 | 0.512** |
| Code 6 | 0.574** |
| Total Course Points | 0.564** |

Note: N = 45, *$p < .05$, **$p < .01$

The results of the correlations analysis reveal that students who reflected more on textbooks, homework, and worksheets tend to be less metacognitive during problem solving activities ($r$ = -0.436, $p$ = 0.03). However, students who reflected more on classroom activities ($r$ = 0.512, $p$ = 0.00) and reflected on their own knowledge ($r$ = 0.574, $p$ = .000) tend to be more metacognitive during problem solving activities. Lastly, students who did well in the course and who are assumed to have a high chemistry content knowledge ($r$ = 0.564, $p$ = 0.00), tend to also be more metacognitive during problem solving activities. Therefore, students who are highly metacognitive in their writing by reflecting on interactions within the classroom (Code 5) and on their current knowledge (Code 6) of the chemistry content tend to be more metacognitive when they are engaged in problem solving of chemistry problems.

## Conclusion

The objective of this study was to determine the effect of journal writing in a natural science course for pre-service elementary education majors. It is not common in science classes for students to be able to write down their learning process of a certain science topic and integrate this to formulate a final understanding. Reaching this point of reflection in their journal helps students self-regulate their learning. The use of journals in the classroom provided students an opportunity to become self-regulators of their learning. These entries made by the students helped facilitate discussion with the instructor that might not have happened in the classroom or laboratory. Having the instructor aware of student thinking can help guide students in the right direction to correctly monitor their understanding. This reflection process is important for the instructor to reflect back on. Cooper *et al* (2008) noted that it is essential for not only the student to be aware of metacognition but also for the instructor to be aware of how students are learning. Students were able to express what they know and what they are struggling with. Some students even wrote about how they would try and better their understanding if they were finding it hard to grasp the central ideas of the topic. In their writing they also talked about where they developed this understanding from for certain topics such as from prior experiences, the lecture, text resources, laboratory, discourse with peers, the teaching assistant, or the instructor.

Our use of specific codes to categorize different types of reflection revealed that the positive relationship of reflective writing and content understanding depends also on the type of reflection. As noted earlier, students with strong content were very likely to write about their current understanding as developed from their own thinking (code 6) or from classroom activities (code 5). Students who reflected on text resources tend to have a lower content understanding of the chemistry topics in the course. Reflecting on learning from the classroom events provides a better opportunity for students to experience new learning situations by engaging in activities with others. When students reflect solely on text resources they are more likely to be engaged in the activity alone or to take a simple reading of the text and use it as the basis of understanding, and this does not, we believe, provide the student with new experiences to help facilitate change in understanding. A possible threat to the validity of this research is that we did not know anything about the writing of the students coming into the course. Knowing the students' prior writing skills would help us determine how well students are at expressing ideas. If students struggle at this task they may struggle with reflecting on their learning and being able to formulate well structured reflections.

## Implications for Research and Instruction

The journal produced by the student is an artifact that can be used at the end of the semester as a planning resource for exams. The student is aware of the understanding they had by referencing the journal and are able to evaluate their current understanding to see if understanding has changed at the end of the term. At this point the student can monitor what they need to review in order to succeed on exams. This whole process is a good source of self-regulation for students to take part in, and gives them time to sit down and think about their understanding of science topics. Also, it is a substantial and consistent tool for students to learn about themselves as learners.

This research confirmed a relationship between content knowledge and journal writing, supporting the idea that this writing-to-learn activity is valuable as a means of supporting students in the exploration and understanding of their learning processes. It has been noted that the reflection process offers an opening for students to understand their learning process and increase their active involvement in learning (Moon, 1999). It also helps students gain a personal ownership of their learning.

These findings add to the research literature on the analysis of writing-to-learn and reflection in instructional learning of science. The findings show that not all types of reflection correlate with improved learning, which implies that it is essential for researcher to examine the content of specific reflective students, not just overall reflection. Second, the link of metacognition to learning in this work is strengthened by the additional use of the MCA-1 survey as an independent measure of metacognition in problem solving.

There are three implications on writing-to-learn as a result of this research study. First, students may benefit from directed reflections (Davis, 2003). This benefit requires activities, and this is an active classroom (Varales et al, 2008). A major component of direct reflection is knowledge growth within the learner (Davis, 2003). Second, caution about what may be rote recapitulation of textbook material. Lastly, if students know there is a potential direct relation between journal topics and assessment, they may perform better on both.

The second major implication relates to the particular benefit of reflection in both writing and in problem solving. McCrindle and Christensen (1995) also found that journal writing tasks in a science class not only promote metacognitive thought but also helped learners gain more developed cognitive strategies and attain better learning outcomes.

Third, having students write reflective journals is a very good way to detect how they express metacognitive awareness, complementing other strategies (Mattox, Reisner, & Rickey, 2006; Greenbowe & Hand, 2005; Cooper, Sandi-Urena, Stevens, 2008; Keys et al, 1999). But the findings of specific benefits from reflection on knowledge (an explicit metacognitive act) and on activities, but not on the textbook, also suggest that providing an activity-rich environment for students to use as the basis of their reflection is important. These results seem to provide additional support for the effectiveness of learning in a mode where the student actively creates knowledge, often in conjunction with other peers (Cooper, 2005).

## Acknolwedgements

This project has been supported by an NSF Course, Curriculum, and Laboratory Improvement Grant, DUE-0311624.

## References

Baker, L. (1991). Metacognition, comprehension monitoring, and the adult reader. *Educational Psychologist*, 1, 3-38.

Bangert-Drowns, R., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, 74, 29-58.

Britton, J. R. (1970). Language and learning. London: Allen Lane.

Boud, R. Keogh, and D. Walker (Eds.), Turing Experience into Learning (pp. 19). EastBrunswick, N.J.: Nichols.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other mysterious mechanisms. In F. E. Weinert and R. H. Kluwe (Eds), Metacognition, Motivation, and Understanding (pp.65-116). Hillside, NJ: Lawrence Erlbaum.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In P. H. Mussen (Ed.), Handbook of child psychology: Cognitive development, (Vol. III, pp. 77-166). New York: Wiley.

Carcolice, M. S. (2005). Introduction to the science writing heuristic. In N. J. Pienta, M. M. Cooper, and T. J. Greenbowe (Eds) Chemists' Guide to Effective Teaching (pp.12-27). Upper Saddle River, NJ: Prentice Hall.

Cooper, M. M. (2005). An introduction to small-group learning. In N. J. Pienta, M. M. Cooper, and T. J. Greenbowe (Eds.), Chemists' Guide to Effective Teaching (pp.140-154). Upper Saddle River, NJ: Prentice-Hall.

Cooper, M. M., Cox, C. T., Nammouz, M., & Case, E. (2008). An assessment of the effect of collaborative groups on students' problem-solving strategies and abilities. *Journal of Chemical Education*, 85, 866-872.

Cooper, M. M., Sandi-Urena, S., & Stevens, R. (2008). Reliable multi method assessment of metacognition use in chemistry problem solving. *Chemistry Education Research and Practice*, 9, 18-24.

Davis, E. A., (2003). Prompting middle school science students for productive reflection: Generic and directed

prompts. *The Journal of the Learning Sciences*, 12, 91-142.
Emig, J. (1977). Writing as a mode of learning. College Composition and Communication, 28, 122-128.
Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive development inquiry. *American Psychologist*, 34, 906-911.
Greenbowe, T. J., & Hand, B. (2005). Introduction to the science writing heuristic. In N. J. Pienta, M. M. Cooper, and T. J. Greenbowe (Eds.), Chemists' Guide to Effective Teaching (pp.140-154). Upper Saddle River, NJ: Prentice-Hall.
Keys, C. W., Hand B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, 36, 1065 1084.
Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, 11, 203-270.
Mason, L., & Buscolo, P. (2000). Writing and conceptual change: What changes? *Instructional Science*, 28, 199 226.
Mattox, A. C., Reisner, B. A., & Rickey, D. (2006). What happens when chemical compounds are added to water? An introduction to the model-observe-reflect-explain (MORE) thinking frame. *Journal of Chemical Education*, 83, 622-624.
McCrindle, A. R., & Christensen, C. A. (1995). The impact of learning journals on metacognitive and cognitive processes and learning performance. *Learning and Instruction*, 5, 167-185.
Moon, J. (1999a). Reflection in Learning and Professional Development (pp.188-194). London: Kogan Page.
Novak, J. D., & Gowin, D. B. (1984). Learning How To Learn. New York, NY: Cambridge University Press.
Puntembekar, S., & Kolodner, J. L. (1998). Distributed scaffolding: Helping students learn in a learning by design™ Environement. *Proceedings International Conference of the Learning Sciences '98*, pp.35-41.
Rickey, D., & Stacy, A. M. (2000). The role of metacognition in learning chemistry. *Journal of Chemical Education*, 77, 915-920.
Varelas, M., Plotnick, R., Wink, D., Fan, Q., & Harris, Y. (2008). Inquiry and connections in integrated science content courses for elementary education majors. *Journal of College Science Teaching*, 37, 40-47.
Wink, D. J. & Choe, J. H. (2008). Pennies and Eggs: Initiation into inquiry learning for preservice elementary education teachers. *Journal of Chemical Education*, 85, 396-398.
Wink, D. J., Ellefson, J., Nishimura, M., Perry, D., Wenzel, S., & Hwang-Choe, J (2009). Fostering pre-service teacher identity through student-initiated reflective projects. *Feminist Teacher*, 19, 151.

# Fostering Online Search Competence and Domain-Specific Knowledge in Inquiry Classrooms: Effects of Continuous and Fading Collaboration Scripts

Christof Wecker, Ingo Kollar, Frank Fischer, University of Munich, Leopoldstr. 13, D-80802 Munich
Helmut Prechtl, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Olshausenstr.62, D-24098 Kiel
christof.wecker@psy.lmu.de, ingo.kollar@psy.lmu.de, frank.fischer@psy.lmu.de, prechtl@ipn.uni-kiel.de

**Abstract:** Collaborative inquiry learning can be regarded as a fruitful approach to foster students' scientific literacy for participation in societal debates involving scientific issues. In a four-week field study in biology with 131 ninth-grade students from six classrooms distributed over three experimental conditions, we investigated the effects of a continuous and a fading collaboration script for collaborative online search compared to unsupported collaboration on students' online search competence and domain-specific knowledge. Findings indicate a superiority of a continuous collaboration script compared to unsupported collaboration with respect to both students' online search competence and domain-specific knowledge, whereas the fading collaboration script had a significant positive effect compared to unsupported collaboration only with respect to students' online search competence. These findings extend our knowledge from laboratory research on effects of collaboration scripts on domain-general competences and domain-specific knowledge.

## Scientific literacy as a goal of science education in schools

A main purpose of science education is not so much to educate future scientists or engineers, but rather to prepare all students as responsible citizens for participation in societal debates that involve scientific issues. Current examples of such issues are the prospects and risks of nuclear power, the genetic engineering of crops and other plants for purposes of food production, or preimplantation genetic diagnosis.

For the development of reasoned positions about these questions, scientific literacy is an important prerequisite (Laugksch, 2000), which involves both fundamental knowledge about important scientific principles and the ability to gather and evaluate more specific and recent information that goes beyond what can ever be learned in school. Current information and communication technology has made such information widely accessible. Accordingly, we regard domain-specific scientific knowledge, i. e. knowledge of basic terms, knowledge of important facts and understanding of fundamental explanatory principles, as well as online search competence, i. e. the ability to localize and evaluate relevant scientific information on the internet, as two crucial components of scientific literacy that all students should develop in order to be able to participate in tomorrow's science-related societal debates. To develop instruments for measuring and support for fostering online search competence, we conducted a cognitive task analysis of internet search tasks which ask learners to use the internet to gather evidence they could use to formulate arguments in societal, science-related debates (see Kollar, Wecker & Fischer, 2009). Following that, we validated the resulting model of competence by asking an online search expert, a content expert and a search novice to think aloud during such tasks. Most importantly, the think aloud study showed that both experts, after having built an initial argument sketch, established a scheme for what counted as evidence to support their position and let this evidence scheme guide their internet search behaviour. The novice, in contrast, browsed through the internet in a rather mindless fashion, not much really taking care of relevance, credibility or scientific quality of web site content, finally arriving at less sound arguments that were supported by less reliable information found on the Web.

## Collaborative inquiry learning as an approach to foster scientific literacy

Scientific literacy and its sub-components domain-specific knowledge and online search competence can be fostered in inquiry learning contexts (e.g., de Jong, 2006), which are implemented with the aid of digital media and internet technologies that afford collaboration between students (see Slotta & Linn, 2000). In web-based inquiry learning, students use computer and internet technologies (usually in groups) to work on authentic scientific phenomena or problems in a way similar to scientists. It can be argued that by engaging in such activities, scientific literacy may be developed, since they typically involve elaborating upon domain-specific scientific knowledge and using digital media to gather evidence, which may foster online search competence.

As empirical evidence shows, however, for inquiry learning to be productive, it needs to be structured and scaffolded appropriately (de Jong & van Joolingen, 1998). Therefore, web-based environments include a variety of scaffolds designed to facilitate students' inquiry processes. For example, the Web-based Inquiry Science Environment (WISE; Slotta & Linn, 2000) includes hint questions designed to trigger elaborations, mapping tools to visualize the relations between multiple variables, or background information on the given science problem. As an analysis of WISE and several other web-based inquiry learning environments shows, however, most environments lack support that directly aims at scaffolding the collaboration process between the learning partners. Against the background of empirical research on collaborative learning, which shows that collaboration rarely reaches high quality when it happens spontaneously (Gillies, 2003), the use of scaffolds directed at an improvement of collaboration seems to be warranted.

### Collaboration scripts to support collaborative inquiry learning

One promising way of providing such support is to develop collaboration scripts (Kollar, Fischer & Hesse, 2006; Dillenbourg & Jerman, 2007) and integrate them into web-based inquiry learning environments. Collaboration scripts are socio-cognitive scaffolds that specify activities, sequence them and distribute them among different roles taken over by members of a small group of learners (Kollar et al., 2006).

Despite some research about the effects of collaboration scripts on learning processes (e. g. De Wever, Van Keer, Schellens & Valcke, 2009), their effects on learning outcomes have been investigated mainly in laboratory experiments. Several lab studies about computer-supported collaborative learning during online discussions in a problem-based learning context demonstrated beneficial effects of different kinds of collaboration scripts on students' domain-general competencies (Weinberger, Ertl, Fischer & Mandl, 2005; Stegmann, Weinberger & Fischer, 2007). However, only for some types of computer-supported collaboration scripts, e.g., a peer-review script, beneficial effects with respect to the domain-specific knowledge could be found (Weinberger et al., 2005; Diziol, Rummel, Spada & Mclaren, 2007). In the context of web-based inquiry learning, a collaboration script aimed at helping students to form well-grounded arguments, counterarguments and integrations was developed and integrated at specific points of as WISE curriculum unit (Kollar, Fischer & Slotta, 2007). In line with previous research, this script helped students develop higher levels of the domain-general competence of argumentation, but did not help them reach higher levels of domain-specific knowledge when compared with less structured collaboration. Moreover, it appeared that during the learning process the collaboration script only raised argumentation quality as long as it was present. As soon as the script was "switched off", learners relapsed into their previous argumentation style. In other words, internalization of the argumentation strategy was low, which may be explained by a rather short learning time of 120 minutes. The question whether computer-supported collaboration scripts produce more robust effects when longer learning phases are studied is yet unanswered.

### The fading of collaboration scripts

The laboratory research mentioned clearly indicates a need for finding ways to support the internalization of strategies targeted by collaboration scripts. One such approach is the fading of support, which is considered as part and parcel of scaffolding (Pea, 2004; Wood, Bruner & Ross, 1976). Up to now research on the effects of the fading is sparse, and the results are mixed. In one of two experiments, evidence for beneficial effects of fading was found, whereas in the other one performance decreased in the process of fading (Leutner, 2000). In a second study a marginally significant positive effect of fading was demonstrated with respect to only one of several aspects of knowledge about the principles of scientific explanations (McNeill, Lizotte, Krajicik & Marx, 2006). No effect of fading on the quality of explanations in the posttest was found in a further study, while during the learning phase the quality of explanations decreased in the process of fading (Lee & Songer, 2004).

An empirical study about the effects of the *fading of scripts* showed that fading alone had no positive effect on learning (Wecker & Fischer, 2007). If it was combined with "distributed monitoring" (cf. King, 1998), i. e. when the learning partner was asked to monitor whether his or her peer complied with the strategy induced by the script, a significant positive effect on students' strategy knowledge in comparison to a continuous script was found. However, this study was also conducted as a lab experiment with a rather short learning time.

### Research questions

The summary of the current state of research provided above warrants further consideration whether the effects of small-group collaboration scripts as well as their fading show up also in real science classrooms. Furthermore, it needs to be investigated whether collaboration scripts as well as their fading can foster also domain-specific knowledge, if more extended learning phases are implemented. The opportunity to shift the focus from the application of a domain-general strategy targeted by the collaboration script to the elaboration of domain-specific content to be learned during longer learning phases would warrant such expectations.

Accordingly, our research questions were the following:

(1) What are the effects of a small-group collaboration script as well as its fading on students' domain-specific knowledge?

We expected that in classroom-based inquiry learning over a comparably extended period of time domain-specific knowledge is fostered if collaboration is supported by a collaboration script compared to collaboration without a collaboration script. According to the line of reasoning just presented, a beneficial effect of a fading collaboration script on students' domain-specific knowledge could also be expected.

(2) What are the effects of a small-group collaboration script as well as its fading on students' online search competence?

In line with the bulk of previous findings about the effects of small-group collaboration scripts from the laboratory we assumed a positive effect of a collaboration script compared to collaboration without a collaboration script on students' online search competence. We further hypothesized that a fading collaboration script would have a beneficial effect on students' online search competence compared to both collaboration without a collaboration script and collaboration supported by a continuous collaboration script.

## Method

### Participants and design

The participants were 131 students from six ninth-grade classes from three urban high schools. They were on average 14.7 years old (*SD* = 0.75); 53 of them were girls, 78 were boys. The decision to participate in the project rested with the biology teachers of the classes. However, data from the individual students were only collected if their parents had agreed individually.

Three experimental conditions differing in the amount of instructional support for collaborative online search were compared in a one-factorial quasi-experimental design with the classes as the units of random assignment. The first condition involved no collaboration script, the second a continuous collaboration script, and the third a fading collaboration script (see table 1).

Table 1: Design of the study.

| *No collaboration script* | *Continuous collaboration script* | *Fading collaboration script* |
|---|---|---|
| N = 43 students<br>from 2 classes | N = 34 students<br>from 2 classes | N = 54 students<br>from 2 classes |

### Curriculum unit and instructional setting

The study was conducted in the context of an inquiry-based curriculum unit that spanned seven regular biology lessons, which were preceded and followed by one test session each (see table 2). During the unit, the students were supposed to arrive at a decision about whether they supported or rejected "green" Genetic Engineering, i. e. the genetic modification of plants for the purpose of food production. All lessons in this curriculum unit were led by the regular biology teachers of the participating classes.

Table 2: Procedure of the study and phases of the instructional unit.

| *Cycle* | *Lesson* | *Activity* | *Duration* |
|---|---|---|---|
| | 0 | Pretest | 45 min |
| | 1 | Introduction to background domain knowledge | 45 min |
| 1 | 2 | Introduction to *economic aspects* of "green" genetic engineering<br>Introduction to the online search strategy<br>Studying of background information relevant to *economic aspects* of "green" genetic engineering | 5 min<br>5 min<br>35 min |
| | 3 | Online search on *economic aspects* of "green" genetic engineering | 45 min |
| | 4 | Discussion about *economic aspects* of "green" genetic engineering | 15 min |
| 2 | | Introduction to *ecological aspects* of "green" genetic engineering<br>Studying of background information relevant to *ecological aspects* of "green" genetic engineering | 5 min<br>20 min |
| | 5 | Online search on *ecological aspects* of "green" genetic engineering<br>Discussion about *ecological aspects* of "green" genetic engineering | 30 min<br>15 min |
| 3 | 6 | Introduction to *health aspects* of "green" genetic engineering<br>Studying of background information relevant to *health aspects* of "green" genetic engineering | 5 min<br>40 min |
| | 7 | Online search on *health aspects* of "green" genetic engineering<br>Discussion about *health aspects* of "green" genetic engineering | 30 min<br>15 min |
| | 8 | Posttest | 45 min |

The unit started with a short introduction to relevant background domain knowledge from Genetics. To keep the information about the strategy of online search to be acquired constant, the students in all three conditions also received the same introduction to this strategy.

Lessons two through seven consisted of three consecutive learning cycles about three different topical aspects of the discussion about "green" Genetic engineering: one on economic, one on ecological, and one on health aspects. Each cycle consisted of three consecutive steps. First, the students were asked to collaboratively browse through an online library in dyads in order to gather background information about Genetics and Genetic Engineering relevant to the current topical aspect of the discussion. In the second step, student dyads conducted collaborative online searches to support, discard or modify their initial argument concerning the current topical aspect of the discussion according to the strategy they were introduced to before. In these phases of collaborative online search, which took place in lessons three, five and seven, the independent variable was manipulated (see below). In the third step, the teacher led a plenary discussion in which students exchanged their elaborated arguments based on the findings from their collaborative online searches.

Students collaborated face-to-face sitting next to each other. The technical equipment both to access the online library and conduct the online searches comprised a wireless network and one laptop computer with a mouse per student. The students kept these computers throughout the time of each lesson and received the same computer each time, which served two purposes: On the one hand, students were enabled to create documents and to revisit or edit them in later sessions; on the other hand, this procedure facilitated data collection in that it was easy to track which student had worked on which computer.

The online library was implemented as a module in the Web-Based Inquiry Environment (WISE). It comprised three pages with assignments for the three topical aspects of the discussion (economic, ecological and health aspects) and six sections about topics from Genetics and Genetic Engineering designed on the basis of regular ninth-grade Biology school books.

During the collaborative online search phases, the browsers of the collaborating students in each dyad were connected via a software tool named S-COL (Wecker et al., 2009). This allowed for collaborative internet browsing, i.e., during their online searches, both learning partners from each dyad always saw the same web pages, no matter who of them clicked on a link or performed a web search.

## Independent variables

The amount of instructional support for collaborative online search was varied among a condition without collaboration script, a condition with a continuous collaboration script, and a condition with a fading collaboration script.

### No collaboration script

In this condition, the browsers of the two students in each dyad were connected as described. These students received no support beyond the teacher's introduction to the strategy of online search at the beginning of the curriculum unit. This information was equivalent to the information presented in the collaboration scripts in the other two conditions.

### Continuous collaboration script

In addition to connecting the computers of the two partners of a dyad, the S-COL software tool described above was used in the two experimental conditions to display particular prompts on the basis of the type of website the students were accessing (Google start page, Google hit list, any other web site).

In each dyad there were two roles (A and B) that switched after returning to Google (which they were required to use for their searches) from any other web page encountered during the search activities. The collaboration script was implemented as complementary text prompts in the scaffolding areas of S-COL in the browsers of both group members (see left part of the screen in figure 1). The script contained prompts for helping students formulate an initial argument and sketch the information needed, select search terms, evaluate the hit list, localize relevant information on a web page, and write the final elaborated argument. For example, during the selection of search terms, learner A was prompted to suggest a set of terms and discuss them with B, while B had the task to first recall the information they had decided to look for, and comment on A's suggestions for the search terms with respect to their likelihood of yielding suitable as well as inappropriate hits.

### Fading collaboration script

In the condition with the fading collaboration script the same prompts as in the condition with the continuous collaboration script were used to support collaborative online search, but they were gradually removed (faded out) over the whole extension of the curriculum unit depending on the number of online searches the students had performed. After a series of four external web sites accessed, the prompts described before became more unspecific: Initially, the scaffolding area of S-COL contained both the names of the individual steps as well as explanatory text. In the second fading phase, only the names of the steps were displayed. In the final phase, only headings for the actual activity were displayed.

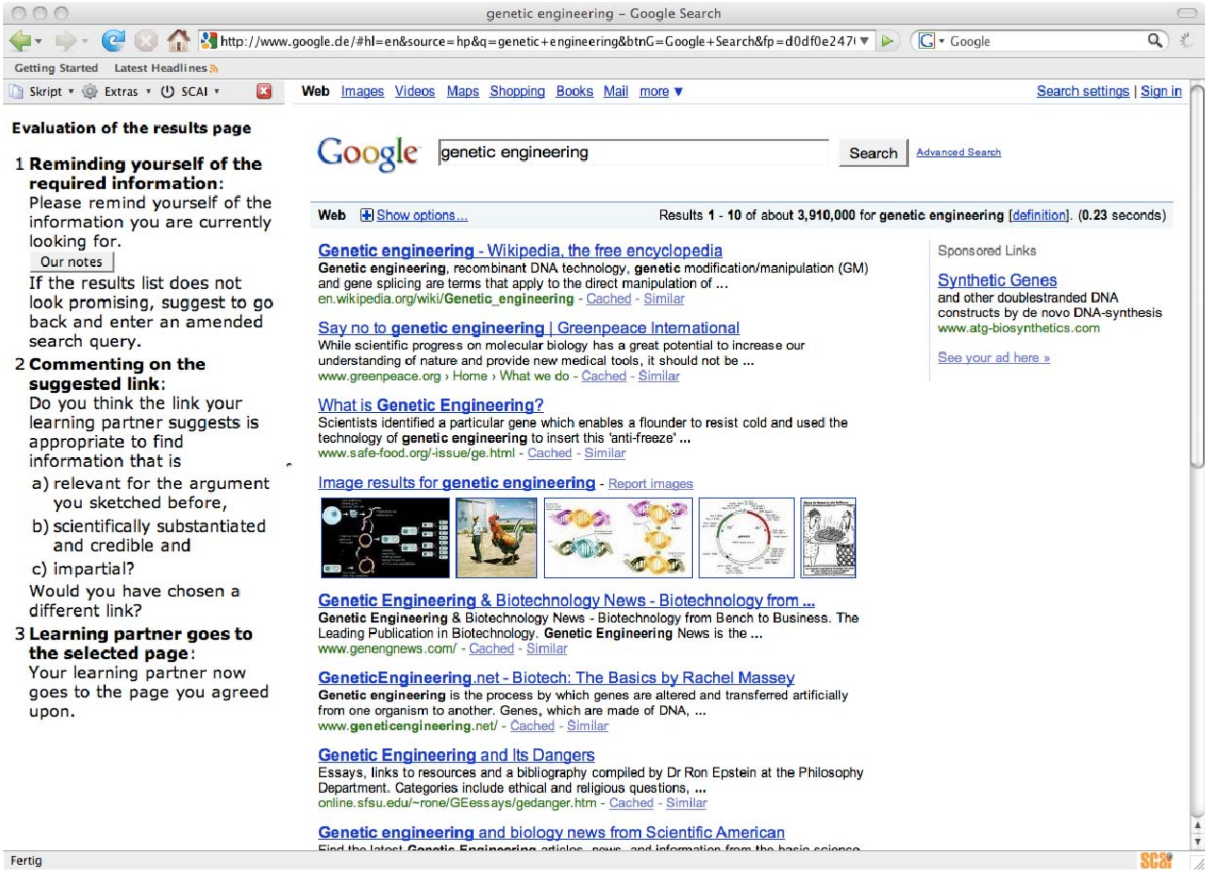



Figure 1: Part of the continuous collaboration script presented along with a Google hit list (collaboration script prompts for one of the two learners displayed in the area on the left; Google hit list displayed in the area on the right)

## Procedure

The procedure of the study was as follows (see table 2): The biology lesson immediately before the start of the instructional unit was used for a 45-minute pre-test for online search competence, domain knowledge and control variables. The following phase was constituted by the seven lessons described above, distributed over a period of three and a half weeks. This phase included the manipulation of the independent variable in lessons three, five and seven. In the biology lesson immediately following the instructional unit, a 45-minute post-test was conducted, which included online search competence, domain knowledge and further control variables.

## Dependent variables

### Domain-specific knowledge

Domain-specific knowledge was measured in the post-test by a test that consisted of 18 items about Genetics and Genetic Engineering. One third of them covered knowledge about concepts, another third factual knowledge and another third understanding of explanatory relations. The items from these three subscales were evenly distributed over different types of answering format (six multiple choice, twelve open). A six-item version of the test with precisely the same structure was used to measure domain-specific knowledge in the pre-test. The answers to the items with an open answering format were analyzed with respect the occurrence of propositions taken from an extremely elaborate expert solution that contained any correct proposition that could conceivably be part of a student's solution For each of these propositions it was coded whether the student's answer contained it or not. Three coders coded 10 % of the material independently of each other to assess the objectivity of the analysis. The average percentage of agreement with respect to the different coding variables was 98 %; Cohen's kappa amounted to $\kappa$ =.94 on average. For each item from the test the proportion of the propositions that were included in the student's answer was determined (values between 0 and 1). These proportions as well as the scores from the multiple-choice items were summed up across the different items, yielding an overall scale for *domain-specific knowledge* ranging from 0 to 18 with a reliability of Cronbach's $\alpha$ = 59). Three subscales ranging from 0 to 6 for *knowledge about concepts*, *factual knowledge* and *understanding* are also used in the statistical analyses. Due to the highly elaborated nature of the underlying expert solution, even low scores indicate substantial knowledge.

### Online search competence

Online search competence was measured by a task in which the students were asked to describe in as much detail as possible how they would use the internet to form a position about a specific sample issue without actually doing so. In the pre-test, the sample issue the students could use for their description was whether mobile phone transmitters should be forbidden in the neighbourhood of day nurseries. In the post-test, the corresponding issue was whether nuclear power plants should be abandoned. The answers were pre-structured by means of a two-column table of up to eight rows. The left-hand column was to be used for the description of the single steps involved in the search, whereas the right-hand column was to be used for the description of the evaluative criteria to be applied in combination with the specific steps. Both pre- and post-test were coded for the occurrence of each individual element of a general expert solution that contained all the steps as well as all the evaluative criteria from the strategy explained to the students in all three conditions in the introduction and suggested by the collaboration script in the conditions with continuous or fading collaboration scripts. The average percentage of agreement with respect to the different coding items was 96 %; Cohen's kappa amounted to $\kappa$ =.65 on average. An overall scale for *online search competence* was formed by counting the positively coded variables, with a reliability of Cronbach's $\alpha$ = .57. Separate scales for *knowledge about single steps* and *knowledge about evaluation criteria* were also formed (with some coding items excluded from both subscales). As in the case of domain-specific knowledge, even low scores indicate considerable amounts of competence due to the complexity of the underlying expert solution.

### Statistical analysis

The significance level was set to 5 % for all analyses. All analyses of variance or covariance were conducted with the manipulated instructional support as one independent factor and the classrooms as a further independent factor nested within the experimental conditions. The covariates used were significantly correlated with the dependent variables.

## Results

### Research question 1: Effects of a small-group collaboration script and its fading on students' domain-specific knowledge

With respect to the effects of the two collaboration script versions on students' domain-specific knowledge, analyses of covariance were conducted with the overall domain-specific knowledge scale and the subscales for knowledge about concepts, factual knowledge and understanding as dependent variables, classes nested within the three kinds of instructional support as well as the kind of instructional support as independent variables, and the overall domain-specific prior knowledge as a covariate. As can be seen from table 3, with respect to all four dependent measures students from the condition with the continuous collaboration script outperformed both students from the condition without a collaboration script and from the condition with the fading collaboration script. A significant main effect of the kind of instructional support was detected for the overall scale of domain-specific knowledge ($F$(2; 124) = 7.66; $p$ < .01; partial $\eta^2$ = .11), which amounts to a medium to large effect. A planned comparison between the condition without a collaboration script and the condition with a continuous collaboration script was significant, $F$(1; 72) = 5.49; $p$ = .02; partial $\eta^2$ = .07. The planned contrast between the condition without a collaboration script and the fading collaboration script did not reach significance in the expected direction, $F$(1; 92) = 2.64; $p$ = .11; partial $\eta^2$ = .03.

Further analyses revealed that the detected effect of the kinds of instructional support is due to parallel patterns of significant effects of approximately medium size in all three subscales of domain-specific knowledge: knowledge about concepts, $F$(2; 124) = 4.54; $p$ = .01; partial $\eta^2$ = .07, factual knowledge, $F$(2; 124) = 3.27; $p$ = .04; partial $\eta^2$ = .05, and understanding, $F$(2; 124) = 4.98; $p$ = .01; partial $\eta^2$ = .07.

With respect to one of the scales of domain-specific knowledge, knowledge about concepts, also substantial between-classroom variance was found ($F$(3; 124) = 4.27; $p$ = .01; partial $\eta^2$ = .09) which may be caused by factors beyond our experimental conditions, e.g. the teacher's behaviour or the class climate.

### Research question 2: Effects of a small-group collaboration script and its fading on students' online search competence

With respect to students' online search competence, analyses of covariance were conducted with the overall online search competence scale and the subscales for knowledge about single steps and knowledge about evaluation criteria as dependent variables, classes nested within the three kinds of instructional support as well as the kind of instructional support as independent variables, and the overall prior online search competence as a covariate. Table 4 shows that descriptively students in the two conditions with the continuous and the fading collaboration script outperformed students in the condition without a collaboration script on the overall online search competence scale as well as the two subscales knowledge about single steps and knowledge about evaluation criteria. The main effect of the kinds of instructional support with respect to the overall scale was significant and of medium size, $F(2; 124) = 5.16$; $p = .01$; partial $\eta^2 = .08$. Both the planned comparison between the continuous collaboration script and no collaboration script, $F(1; 72) = 9.97$; $p < .01$; partial $\eta^2 = .12$, and the planned comparison between the fading collaboration script and no collaboration script, $F(1; 92) = 7.17$; $p = .01$; partial $\eta^2 = .07$, were significant. However, the final planned comparison between the fading collaboration script and the continuous collaboration script was not significant, $F(1; 83) < 1$; $p = .81$; partial $\eta^2 < .01$.

Table 3: Means and standard deviations of the dependent variables in the three experimental conditions.

| | | *No collaboration script* | | *Continuous collaboration script* | | *Fading collaboration script* | |
|---|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| **Domain-specific knowledge** | **Pretest** | **0.67** | **0.72** | **0.68** | **0.68** | **0.78** | **0.69** |
| | **Posttest** | **3.58** | **1.53** | **4.54** | **1.81** | **3.14** | **1.88** |
| **Knowledge about concepts** | **Pretest** | **0.14** | **0.23** | **0.07** | **0.18** | **0.19** | **0.28** |
| | **Posttest** | **1.49** | **0.77** | **1.95** | **0.93** | **1.40** | **1.00** |
| **Factual knowledge** | **Pretest** | **0.12** | **0.32** | **0.09** | **0.29** | **0.11** | **0.32** |
| | **Posttest** | **0.90** | **0.79** | **1.26** | **0.68** | **0.90** | **0.75** |
| **Understanding** | **Pretest** | **0.42** | **0.49** | **0.52** | **0.54** | **0.47** | **0.54** |
| | **Posttest** | **1.19** | **0.84** | **1.33** | **0.74** | **0.84** | **0.76** |
| **Online search competence** | **Pretest** | **3.46** | **2.10** | **3.03** | **2.43** | **3.59** | **2.33** |
| | **Posttest** | **2.65** | **2.27** | **3.68** | **2.41** | **3.85** | **2.65** |
| **Knowledge about single steps** | **Pretest** | **2.35** | **1.48** | **2.06** | **1.76** | **2.39** | **1.77** |
| | **Posttest** | **1.74** | **1.48** | **2.56** | **1.66** | **2.61** | **1.94** |
| **Knowledge about evaluation criteria** | **Pretest** | **1.00** | **0.82** | **0.85** | **0.86** | **0.96** | **0.82** |
| | **Posttest** | **0.86** | **1.08** | **1.03** | **1.11** | **1.06** | **1.04** |

Further analyses demonstrated that the significant effect with respect to the overall scale is due only to an effect of the kinds of instructional support on knowledge about single steps, $F(2; 124) = 5.49$; $p = .01$; partial $\eta^2 = .08$. There is no significant corresponding effect with respect to knowledge about evaluation criteria, $F(2; 124) < 1$; $p = .41$; partial $\eta^2 = .01$.

With respect to all scales of online search competence, i. e. the overall scale of online search competence ($F(3; 124) = 4.19$; $p = .01$; partial $\eta^2 = .09$) as well as the subscales of knowledge about single steps ($F(3; 124) = 2.96$; $p = .04$; partial $\eta^2 = .07$) and knowledge about evaluation criteria ($F(3; 124) = 3.42$; $p = .02$; partial $\eta^2 = .08$), however, there was also substantial between-classroom variance corresponding to medium effect sizes.

## Discussion

The results indicate that the collaboration script approach (Kollar et al., 2006) could successfully be transferred to structure web-based collaborative inquiry learning in real secondary school classrooms. Indeed, the collaboration script was powerful enough to lead learners to higher levels of both domain-specific knowledge and online search competence than unsupported collaboration. Probably, the script helped learners to engage in higher-level internet search processes and also to show a deeper elaboration of the domain-specific information provided in the online project library. Possibly, the high-level evaluation of information prompted by the collaboration script raised the learners' awareness of the need to produce arguments of a high domain-specific quality so that they could stand during the plenary discussions, which may have contributed to a deeper elaboration of domain-specific information that could be found online.

With respect to the effectiveness of the fading of the collaboration script, however, results are mixed: It did not add to the positive effects of the continuous script on students' online search competence, and did not even lead to better results with respect to domain-specific knowledge than the unstructured control group. Obviously, fading alone does not seem to be enough to foster the internalization of a search strategy but very likely needs to be combined with further instructional support (cf. Wecker & Fischer, 2007). For example, it might be helpful to have the teacher from time to time model crucial steps of the search process in front of the whole class (see Rosenshine & Meister, 1994), so that the students are reminded of the strategy the script presented before it is faded (which we are currently investigating in a follow-up study).

Finally, it deserves mention that on most dependent measures we found considerable variance between classes. Thus, further aspects of the learning situation (such as teaching style or class climate) definitely have an impact on the success or failure of collaboration scripts in web-based collaborative inquiry learning. Future research should clarify what classroom variables influence students' learning outcomes.

## References


de Jong, T. (2006). Technological advances in inquiry learning. *Science, 312*, 532-533

de Jong, T. & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68 (2)*, 179-201.

De Wever, B., Van Keer, H., Schellens, T. & Valcke, M. (2009). Structuring asynchronous discussion groups: the impact of role assignment and self-assessment on students' levels of knowledge construction through social negotiation. *Journal of Computer Assisted Learning, 25*, 177-188.

Dillenbourg, P. & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, H. Mandl, J. M. Haake & I. Kollar (Eds.), *Scripting computer supported communication of knowledge: Cognitive, computational and educational perspectives* (pp. 275-301). New York: Springer.

Diziol, D., Rummel, N., Spada, H. & McLaren, B. (2007). Promoting Learning in Mathematics: Script Support for Collaborative Problem Solving with the Cognitive Tutor Algebra. In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *Mice, Minds, and Society: CSCL 2007* (pp. 39-41). ISLS.

Gillies, R. (2003). The behaviours, interactions, and perceptions of junior high school students during small-group learning. *Journal of Educational Psychology, 95*, 137-147.

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review, 10*, 57-74.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Computer-supported collaboration scripts – a conceptual analysis. *Educational Psychology Review, 18*(2), 159-185.

Kollar, I., Fischer, F. & Slotta, J.D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*, 708-721.

Kollar, I., Wecker, C. & Fischer, F. (2009). *Supporting high-school students' scientific literacy through structured inquiry - a use-inspired basic research program.* Paper presented at the 13th Biennial Conference of the European Association of Learning and Instruction, August 25 – 29, Amsterdam.

Laugksch, R. C. (2000). Scientific Literacy: A conceptual overview. *Science Education, 84*, 71-94.

Lee, H. S. & Songer, N. B. (2004). Expanding an understanding of scaffolding theory using an inquiry-fostering science program. Unpublished manuscript.

Leutner, D. (2000). Double-fading support – a training approach to complex software systems. *Journal of Computer Assisted Learning, 16*, 347-357.

McNeill, K. L., Lizotte, D. J., Krajcik, J. & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences, 15(2)*, 153-191.

Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13(3)*, 423-451.

Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*(4), 479-530.

Slotta, J. D. & Linn, M. C. (2000). How do students make sense of internet resources in the science classroom? In M. J. Jacobson & R. Kozma (Eds.), *Learning the sciences of the 21st century*. Mahwah, NJ: Erlbaum.

Stegmann, K., Weinberger, A. & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2(4)*, 421-447.

Wecker, C. & Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: The role of distributed monitoring. In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *Mice, Minds, and Society: CSCL 2007* (pp. 763-771). ISLS.

Wecker, C., Stegmann, K., Bernstein, F., Huber, M. J., Kalus, G., Rathmeyer, S., Kollar, I. & Fischer, F. (2009). Sustainable script and scaffold development for collaboration on varying web content: The S-COL technological approach. In C. O'Malley, D. Suthers, P. Reimann & A. Dimitracopoulou (Eds.), *Computer supported collaborative learning practices: CSCL2009 Conference Proceedings* (pp. 512-516). ISLS.

Weinberger, A., Ertl, B., Fischer, F. & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33(1)*, 1-30.

Wood, D. J., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.


## Acknowledgments

This research was funded by Deutsche Forschungsgemeinschaft (DFG).

# Disciplinary Knowledge, Identity, and Navigation: The Contributions of Portfolio Construction

Jennifer Turns, Brook Sattler
Human Centered Design & Engineering
jturns,brooks2@u.washington.edu

Deborah Kilgore
Center for Engineering Learning and Teaching, University of Washington
kilgore@u.washington.edu

**Abstract:** In this paper, we look at how the construction of two types of professional portfolios supports engineering student learning. To frame "engineering learning," we use the dimensions brought together by Stevens and his colleagues: disciplinary knowledge, identity, and navigation. We present data from a comparative study in which students constructed one of two types of professional portfolios and provided data through extensive questionnaires. In this analysis, we look at their answers to the first question on the questionnaire ("What did you take away from this experience") in terms of the extent to which the students reported insights about disciplinary knowledge, identity, and navigation as a result of portfolio construction activities and the nature of the insights reported.

## Introduction

As with other forms of professional education, the goals for engineering education are complex and the lists of outcomes or competencies are extensive (e.g., ABET learning outcomes, conclusions from national policy reports such as ENGR 2020, Rising above the Gathering Storm, Grand Challenges). To meet the challenges of a dynamic, rapidly changing global milieu, the engineering education community has been exploring various pedagogical innovations such as problem-based learning and integrated first-year curricula, to provide students exposure to the kinds of complex, multifaceted learning experiences that more closely resemble situations they are expected to face as engineering professionals. Portfolio construction is an activity that has been advocated as valuable in supporting reflection on these and other learning experiences. While the body of work on the educational benefits of portfolio construction has been growing in other disciplines, work remains to be done to help fully understand how portfolio construction can enhance engineering learning.

In this paper, we present data from two portfolio interventions requiring different amounts of time and resources on the part of students and educators, to determine the engineering learning that might be possible given investment in such activities. To examine student learning from a perspective that complements the complex challenges of the present age to engineering education, we used a three dimensional approach proposed by Stevens and his colleagues, to include accountable disciplinary knowledge, identity, and navigation. We found this to be a useful framework for understanding what students perceived to be important benefits of portfolio development.

## Portfolios in engineering education

Portfolios are frequently understood as a collection of student work with a purpose. Among the many reasons driving interest in portfolios are the following: student knowledge is represented by their own work, students can take control of gauging the quality of the work, and the processes of selecting and explaining work in a portfolio can create opportunities for reflection.

Portfolios are not new to engineering education. The various portfolio projects being reported represent goals that are diverse, implementations that range from course to institution-wide, and variation in the extent to which reflection is emphasized. For example, Eris (2007) explored the potential of portfolios to capture significant types of design knowledge, Williams (2002) focused on a systematic process for setting up a portfolio framework that is oriented toward a particular competency in engineering education and supports reflection, Campbell and Schmidt (2005) described an institution level initiative that goes beyond assessment toward career planning as a goal and has reflection as a key feature, and Knott and colleagues (2004) reported on multiple university-level portfolio initiatives along with data on student experiences in these initiatives. Such a snapshot suggests the level of interest in portfolios and the range of uses. Barrett (2007) provides useful ways of laying these uses. For example, she differentiates the idea of using portfolios for assessment *of* learning or assessment *for* learning, and she draws distinctions between a positivistic approach for using portfolios (i.e., providing documentation of true knowledge) and a constructivist approach (i.e., providing a venue for helping students construct understanding). It is not always easy to determine which paradigm is prominent in the portfolio uses in engineering education.

In our work, we are interested in what happens when creation of the portfolio is framed as construction of an argument about the ways in which one is prepared to contribute to future engineering practice, specifically the realm of engineering practice that is of personal interest. Operationally, we have students instantiate this argument in the form of a portfolio containing a professional statement and multiple annotated artifacts. We explain the professional statement as a written narrative in which students explain the ways in which they are prepared to function as an engineer and which functions as the core of the argument. Artifacts are products and by-products of experiences that support the claims made in the statement. Annotations are explanations of artifacts, the experience that gave rise to them, and the link from the artifact to one or more claims.

There are clearly many ways to think of what such a task means theoretically. For example, seen from a lens of situated learning and legitimate peripheral participation (Lave and Wenger, 1991), this notion of portfolio construction invites students to frame or reframe prior experiences as legitimate peripheral participation in the practices of engineering (e.g., linking prior communication experiences in any setting to the kinds of communication that will be required in engineering practice). Clearly such a task is likely to be easier for some experiences than others. For example, an internship at an engineering company or participation in a capstone design project is likely to have much similarity to engineering practice, and thus the portfolio activity may simply involve being able to articulate the ways in which the experience contributed to being prepared. For other experiences, such as being in a leadership role in a sorority or spending a quarter in a music appreciation class, the challenge involves better *understanding* as well as articulating the possible ways in which the experience contributed to preparation for engineering practice. While much work on education is moving toward having students have educational experience that are more “like” engineering practice, it is still important to remember that students are having lots of experiences (some we control, some we do not)—experiences that can be contributing to their preparation for engineering practice and that giving students opportunity to frame the experiences as preparation has the potential to have educational significance.

In our studies, we have been exploring such questions of educational significance as well as questions of educational feasibility (i.e., what resources are required) in relation to different instantiations of this overall idea. Different instantiations involve different ways of constraining the task and different ways of providing students with support. For example, one way of constraining or not constraining the task involves deciding the scope of experiences across which students are invited to base their argument about their preparation. This has led us to explore *cross-curricular engineering portfolios* (CCP) in which students reflect on all of their experiences to date as well as *course-based engineering portfolios* (CBP) in which students focus on how the experiences associated with a specific course have contributed to being prepared. Our program of research has focused on student construction of these two types of portfolios (e.g., Turns, Cuddihy and Guan, submitted; Eliot and Turns, accepted; Guan and Turns, 2007; Lappenbusch and Turns, 2005).

These variations clearly differ in terms of educational feasibility. For example, course-based engineering portfolio activities could be added as a required assignment to any existing course, thereby not requiring significant change to the curriculum. Incorporating a cross-curricular engineering portfolio activity into the student educational experience is not as straightforward. In addition, the larger scope of the cross-curricular engineering portfolio suggests a need for more resources. As a result of such issues, over the past six years, we have studied students constructing these types of portfolios within as well as outside classroom settings

The differential educational significance of these two framings of the portfolio is also open question. Across our studies, we have collected data via observations, interviews, and surveys and our analyses of the data have used a variety of lenses in order to fully understand the role that this intervention can have. For example, we have analyzed our data from a writing-to-learn perspective, a community of practice perspective, an epistemic space perspective, and an identity perspective. Across this work, we have consistently noted that students engaged in constructing these types of portfolios have significant educational experiences and an effort to fully understand these experiences would require a broad framework. The contribution of this paper is to apply such a broad framework to data from our newest study, a comparative study in which student participants constructed one of the two types of portfolios and then told us about their experience through an extensive, open-ended survey.

## Disciplinary knowledge, Identity, and Navigation

In a recent paper, Stevens and his colleagues argued for the value of looking at three dimensions of engineering learning when thinking about the challenges of helping students become engineers (Stevens et al., 2008). These three dimensions--disciplinary knowledge, identity, and navigation--individually and in combination present a rather comprehensive view of how students conceptualize engineering knowledge and themselves as engineers, as they move into, through, and beyond their undergraduate engineering education. Stevens *et al.* present this three-dimensional model as a means for examining the experiences and development of engineering students from the analyst's perspective and illustrate the model with insights developed from ethnographic research. As we will show, the three dimensions may also provide a framework for students to reflect on and characterize *their own* experiences and development.

As Stevens and his colleagues point out, much work on disciplinary education has focused on characterizing and comparing the knowledge of disciplinary experts and student novices. In engineering, for example, researchers have focused on the critical role of design activity to engineering practice and have explored issues such as how expert engineers engage in design, how students engage in design, and how to help engineering students travel the path toward design expertise.

Stevens *et al.* take a significant turn in examining disciplinary knowledge as malleable and situational, rather than the relatively stable and progressively acquired body of knowledge assumed by expert-novice comparisons. They propose the concept of accountable disciplinary knowledge to highlight the idea that while disciplinary knowledge might be defined relative to the profession, the knowledge that is considered "accountable" or relevant changes over the course of a student's academic career. For example, while lifelong learning is considered critical to effective functioning of a professional engineer (i.e., see ABET accreditation criteria), students may not be held accountable to such a notion in the early years of their undergraduate education (or throughout their undergraduate years, for that matter) in favor of bodies of knowledge, like physics or even engineering design, that are more established in the literature and therefore easier to assess. Stevens *et al.* offer the example of engineering education in a particular large public institution that was the subject of their in-depth ethnographic inquiry. The character of knowledge for which students at this institution were accountable changed over the course of their years in college. In their first two years or so, students were assessed largely on their ability to solve discrete closed-ended textbook problems on their own, while in their upper-class years, students were assessed largely on their ability to tackle open-ended ambiguous problems while working in teams. The consequence of such a structural shift in what "counts" as knowledge, not unusual in U.S. engineering programs, is illustrated by the second-year student who found it difficult to reconcile his experiences in a research lab with what he believed would be important for his application to an engineering department.

Influenced by situated learning theory, Stevens *et al.* also assert the importance of identity in understanding engineering learning. Here, identity is not only how a person thinks of herself as an engineer, but also how a person is identified with engineering by others. Thus, identity development is an interactive process involving an individual's efforts to develop a conception of self as an engineer, and the larger community's growing recognition of the individual as an engineer (or not an engineer). As students develop a sense of what does or does not count as engineering, they develop a sense of self as more or less "engineer-like" according to how much or little they find themselves doing what they think counts as engineering. Likewise, others around them identify students as engineers more or less depending on the extent to which the student is doing what others perceive to count as engineering.

This two-sided process is illustrated by Stevens *et al.* in their observations about students at the institution in which they conducted their ethnography. For example, gaining increasing access to limited educational resources served to support the students' growing identities as engineers; while on the other hand, the competitive admissions process at the university (students have to apply to an engineering department during their sophomore or junior year) impeded students' identification as engineers, as they tended to refrain from committing to engineering until a department committed to them. One consequence of such structural conditions is that individuals' identification with the profession may weaken as they worry about admissions.

The third element of the three dimensional view, navigation "focuses on how a person moves through the personal and institutional pathways as an engineer-in-the-making to be officially recognized in one or more ways as 'an engineer,' pathways that are cut along both official and unofficial routes" (Stevens et al., 2008, pp. 356). Navigation provides "a way to talk about how other people and institutional resources play a role in a particular individual becoming an engineer or not" (Stevens et al., 2008, pp. 356). Navigation would include obligatory passage points, official routes, unofficial routes, and detours. For example, an obligatory passage point might be a particularly difficult pre-requisite course and an official route might include taking the required courses in an engineering department. At the same time, there are many unofficial pathways and detours, including internships, summer jobs, extracurricular activities, that can be the greatest sources of variation in students' experiences, engineering learning, and development. Such navigational flexibility can expand the possibilities for a student. For example, working at an engineering job gave one student access to a number of individuals who provided mentoring and additional opportunities for him, not necessarily available to other students.

### Something that happens? Something you can control?

We agree with Stevens and his colleagues that these three dimensions hold great promise for thinking broadly about how engineering students become engineers. What we also wonder is the extent to which these dimensions can be used to characterize the ways in which an intervention helps students take control of their own development. The idea implicit in such a question is that these elements are not only significant for characterizing what has happened to a student along the way, but can also be variables over which a student may have some measure of control. How effectively can students think broadly about the knowledge that they possess and recognize that subsets of their knowledge are accountable for certain situations? To what extent can thinking explicitly about such an issue actually affect the state of disciplinary knowledge? To what extent can trying to explain one's preparedness to be an engineer to someone else affect one's own identification with engineering? To what extent can students effectively monitor their own pathways, making sense of where they have been and adjusting their direction as they move forward? Our experience suggests that students can have agency in these matters, and that this agency can be supported through professional portfolio development activities.

## Approach

In our study, students created either a course-based professional portfolio (CBP) or a cross-curricular professional portfolio (CCP) and shared their portfolio construction experience with us via an extensive online survey. For this paper, we used the three dimensional framework introduced above as a basis for analyzing their response to the first and broadest question on the survey, "what did you take away from this experience?"

### Procedure

All participants attended workshops where they were given instructions and support for creating their portfolios. Participants in the cross-curricular portfolio intervention group attended four 90-minute workshops during the fall of 2008, while participants in the course-based professional portfolio intervention group attended two 90-minute workshops during the winter of 2009. In both cases, workshops were devoted to instructions for creating a professional portfolio and opportunities to peer review one another's work. In the course-based portfolio condition, because of the timing, participants only had the opportunity to peer review the final portfolio. In contrast, participants in the cross-curricular portfolio condition had the opportunity to peer review elements of the portfolio, like the professional statement or specific annotations, as they were developed, as well as the completed portfolio as a whole. Admittedly, these two conditions differed not only in the type of portfolio being constructed, but also the "time on task" and the extent of peer review feedback. The decision to configure the two conditions in this way was based on what was deemed a minimum scale implementation of each type of portfolio as part of an effort to help the engineering community understand what could be possible with these interventions.

A single individual served as the facilitator for all sessions. In addition, to ensure consistency across the sessions, the facilitator's role in these workshops was highly scripted. In particular, the facilitator offered little or no information about the criteria for choosing portfolio content or possible future uses for the portfolio itself. Participants often asked such questions in the workshop and in these cases the facilitator was instructed to turn such questions over to the group for general discussion.

### Participants

A total of 69 undergraduate engineering students participated in the study, 37 in the cross-curricular condition and 32 in the course-based condition. Collectively, the participants ranged in age from 18 to 44 years, with a median age of 21 years. There were approximately the same number of females and males in the study. Because of the relatively small number of participants, we chose to ask students whether they considered themselves underrepresented and over 50% said yes. Participants indicated that they were from a variety of engineering disciplines.

### Data collection

The primary data collected in this study was in the form of extensive online questionnaires that elicited information about the students' experiences of constructing their portfolios. These questionnaires were administered during the final workshop sessions to manage the quality of the data collection. Students spent between 35 and 45 minutes completing the survey. In this exploratory paper, we focus on analyzing the students' responses to the first question—"What did you take away from this experience?" We chose this "take away" formulation rather than a formulation that asked directly about learning (e.g., "what did you learn") for a variety of reasons. Our goal with the first question was to get a sense of students' learning broadly. However, recent work on personal epistemology has helped point out that students themselves have ideas about what counts as knowledge. On our part, we have seen evidence of such issues in other students, when, for example, students described how a portfolio construction activity helped them become more aware of what they know and helped them to see the connections among their different courses and learning experiences, yet subsequently said that they "hadn't really learned anything." In our choice of asking about take-aways, we wanted to try to bypass potentially constrained ideas about what counts as knowledge and learning.

## Data analysis

We used a combination of deductive and inductive analysis of the student "take-away" responses to explore a) the extent to which students reported insights related to disciplinary knowledge, identity, and navigation, b) the nature of these insights, and c) how insights reported by students differed across the two portfolio interventions.

We are using the term deductive analysis to refer to the process of filtering the data relative to each of the three dimensions from the framework introduced earlier. In addition, because initial inspection of the data suggested the prevalence of comments directly related to the portfolio, we added a fourth coding dimension.

- Disciplinary knowledge: a) literal references to "skills" or "knowledge," b) references to other terms that can be considered surrogates for skills and knowledge such as "qualifications" and "qualities"
- Identity: a) literal references to being "an engineer" either in their own eyes or in someone else's eyes, b) references to personal sense of self and "who they are"
- Navigation: a) literal references to "experience," b) references to specific types of experiences such as projects, courses, etc., c) references to accomplishments and achievements because such terms typically refer to experience delineated by a superior quality
- Portfolio: a) references to some aspect of the portfolio experience such as having the portfolio or the process of constructing the portfolio

The two first authors independently coded each student response as related or unrelated to each dimension based on the guidelines presented above. A single response could be coded for multiple dimensions. Reliability was high as measured by Cohen's kappa (0.77 for Disciplinary knowledge, 0.91 for Identity, 0.68 for Navigation, and 0.94 for Portfolio). The lowest value—the value for Navigation—was the result of one coder including mentions of specific courses as a part of navigation. This broader view was adopted for the final coding, and the remaining disagreements were resolved by the first author. Since we were interested in comparing the prevalence of these insights across the two conditions, the coding was done blind to the condition so that preconceptions about the conditions could not influence the coding outcomes.

We used an inductive analysis to characterize the nature of the insights associated with each dimension. The goal of this analysis was to identity themes in the responses coded as related to each dimension. In the results section below, we describe these themes and illustrate the themes using the students' response.

## Results

As depicted in Table 1, the Navigation and Portfolio dimensions were similarly prevalent in the data—both appeared in around 65% of the *overall* set of responses. The dimensions of Identity and Disciplinary Knowledge were less prevalent but still prevalent to a notable degree (i.e., 22% and 30% respectively) in the *overall* responses. In terms of differences between conditions, the prevalence was similar for Disciplinary Knowledge and Portfolio. The CCP condition had a much higher prevalence of Identity and Navigation, while the CBP condition had a much higher prevalence of portfolio-only responses—responses that did *not* address any of the three primary areas of interest (i.e., disciplinary knowledge, identity, navigation). In the remainder of the results, we describe the themes present in responses coded within each dimension. We start with disciplinary knowledge as a category of interest that was not different across conditions, then discuss navigation since it was present but different across the conditions, and then turn to identity which was different across the conditions and almost absent in the CBP condition. We end with comments on the portfolio themes.

Table 1: The percentage (and counts) of the responses addressing each dimension (i.e., extent)

| | Overall (n=69) | CCP (n=37) | CBP (n=32) | Comparison |
|---|---|---|---|---|
| Disciplinary knowledge | 30% (21) | 27% (10) | 34% (11) | CCP < CBP |
| Identity | 22% (15) | 32% (12) | 9% (3) | CCP >> CBP |
| Navigation | 62% (43) | 68% (25) | 56% (18) | CCP >> CBP |
| Portfolio | 65% (45) | 68% (25) | 63% (20) | CCP ~ CBP |
| ---Portfolio Only | 30% (21) | 19% (7) | 44% (14) | CCP << CBP |

## Disciplinary Knowledge

Three themes emerged from inductive analysis of the twenty one responses coded as related to disciplinary knowledge. First, the students' responses showed marked attention to the issue of "accountability" of their knowledge. Most frequently this took the form of identifying or thinking about the portion of their knowledge most relevant to employers or recruiters, but some students also thought about the portion of their knowledge relevant to "professional practice," to "engineering," and to their "engineering career."

- A sense of what skills and valuable experiences I have that I can bring up when talking to recruiters, interviewers, etc. to make myself a stronger candidate; A realization that these experiences/skills that make me a strong candidate are not limited to technical experiences/skills. Examples from other academic disciplines and nonacademic pursuits are also valuable, as are examples related to my major that show nontechnical skills (ability to work in a group, time management, etc.)
- I feel like I can look at a course differently now. What I mean by that is that I can look at a current class and ask myself "what marketable skills am I gaining from this class."

Second, the student responses also suggested that the activity of constructing the portfolio helped students to gain a better sense of the body of knowledge that they had been able to develop by this point in time. In some cases, this was characterized as an inventorying process, while in other cases there was an element of discovery. Such awareness creates a basis for thinking about what knowledge might be accountable to some group.

- However, my chief take-aways from this experience is my realization that I have actually produced a lot more ""artifacts"" than I had remembered. This allowed me to go back and look at the skills I've managed to develop over time. …. (Inventory with a hint of discovery)
- "Building an engineering portfolio helped me to view my completed coursework from a professional perspective as opposed to an academic one. It gave me the opportunity to rethink what I learned and added to the value of the material covered in the course… " (Discovery)

Third, students also reported insights concerning the process for doing the first two activities. Such insights included recognizing the scope of accountable knowledge and how a course can be used to demonstrate knowledge, and even a recognition of the importance taking the time to do such thinking.

- A realization that these experiences/skills that make me a strong candidate are not limited to technical experiences/skills. Examples from other academic disciplines and nonacademic pursuits are also valuable,
- I feel like I can look at a course differently now. What I mean by that is that I can look at a current class and ask myself "what marketable skills am I gaining from this class."
- I haven't really taken the time to think about what it is to be an engineer and show how I have those skills

## Navigation

In the 43 responses that were coded as related to navigation across the two conditions, students 1) revisited experiences they had already had, 2) expressed that they had developed a greater appreciation for their experiences, 3) articulated a broader understanding of the experiences that are relevant to navigation through engineering, and 4) discussed plans for directing future paths. Unsurprisingly, the CBP responses were more likely to emphasize course-specific experiences, although they frequently talked about courses in generally rather than simply the course on which they had focused the portfolio, while the CCP responses emphasized experience more generally.

Several students described how the portfolio development process encouraged them to revisit their prior experiences. For one student, portfolio development simply "forced me to review all of my experiences" while another student "gained a deeper understanding of my professional experience." In addition, students said they developed greater appreciation for the experiences they had had. Some students expressed this greater appreciation in terms of realizing they had accomplished more than they thought, as with one student who wrote, "My chief take-away from this experience is my realization that I have actually produced a lot more "artifacts" than I had remembered." Another student "gained more insight into what I could prove from experiences…. I realized they showed more about me than I had realized before."

Students also described how their understanding of the relevance of their many experiences had expanded, "how everything that I have done up 'til now has some kind of value in the engineering field." This broader conception of relevant experiences, coupled with greater appreciation for such experiences, led students to consider their value: "all my experiences could potentially be documented for future benefits." In looking back at their experiences, students also were compelled to look forward, as one student who planned to continue reflecting on her learning experiences said, "I need to think about each project or assignment as something that is helping me to become a better engineer." In addition to those students who planned to continue reflecting in the future, some students planned to take greater control of their future paths. One student wrote, "I discovered that I needed more to show that I was an engineer. I will now be looking more closely at internships…" In addition to "thinking about everything I have accomplished," some students are also inclined to look forward, at "things I wish to accomplish by the time I graduate…"

Navigation, described by Stevens *et al.* as the dimension that may introduce a great deal of the variation among students' experiences, learning and development, here is shown to present opportunities for students to exert greater control over their learning. Some students say that the portfolio development activity has enabled them to re-think and organize their prior pathways, and be more intentional about those pathways they follow in the future.

## Identity

Since most of the identity-related responses were associated with the CCP condition, a comparative analysis is not strongly warranted. It is possible to note, however, that the three CBP identity-related responses were quite different in emphasis and covered much of the space of the twelve identity-related CCP responses. Of the fifteen responses that were coded as related to identity, most (12) were associated with the CCP condition. Two major themes emerged from inductive analysis of the fifteen responses coded as related to identity. These themes were 1) a changed personal sense of identification with engineering and 2) a changed sense in the ability to explain one's preparedness for engineering. In the process of identifying oneself with engineering, as one student wrote, "I've had to go back and evaluate myself and what I have done as an engineer," some students gained greater identification as an engineer. This same student added, "What I'm mostly going to take out of this is what I have discovered about myself and how everything that I have done up till now has some kind of value in the engineering field." Another student wrote simply that portfolio development, "helped me define who I am as an engineer."

Furthermore, students recognized the importance of being able to explain or prove their preparedness for engineering, as one student wrote, "I have learned how to think about the things I have done in my classes in a way to present myself as an engineer. I learned how to draw on my experiences to come up with specific proofs of my qualities." Another student commented on the value of "having some basis/artifacts for the claims that I am a good engineer." The process also caused students to plan future action toward developing the ability to demonstrate their preparedness. Through reflection on their experiences, some students felt, as one expressed, "I discovered that I needed more to show that I was an engineer. I will now be looking more closely at internships and will be saving assignments." Another recognized the need to pursue "more activity that related to my career as an engineer."

Students' responses also illustrated the two-sided process of identity development as described by Stevens *et al.*, with the individual developing a sense of self as an engineer at the same time and in interaction with others forming a perception of that individual as an engineer. For example, one student wrote, "Sharing with others my experiences and their experiences helps me understand where I stand relative to other engineers in my community." Another also found value in comparisons across other engineers, "I got to look back what I have done as an engineer, and how it may differ from other engineers."

We may assume that the two-sided process of identity development described by Stevens *et al.* is always in play as students interact with their environment, and therefore it is no surprise that the process continues through portfolio development as an educational intervention. The difference here is the possibility for students to look upon their own developing professional identities as the subject of reflection, drawing connections between what they do and who they are becoming. By making sense of these relationships, students gain awareness of the control they can have over their pathways, and thus become more deliberate in developing and claiming their professional engineering identity.

## Portfolio

Two major themes emerged from inductive analysis of the forty-five responses coded as related to portfolio. These themes were 1) having the portfolio and 2) having a process for creating a portfolio. First, students drew attention to the simple fact that they gained from the experience a physical portfolio. Some students went even further and described the portfolio as an organizing tool, and sometimes even specifically to share with others (e.g., "A portfolio that I can present to future employers and academic admissions boards"). The students also students reported gaining knowledge about the "fundamentals of how to build a portfolio" and a sense of how creating a portfolio is related to other types of writing they have learned about (e.g., "Creating a portfolio is quite similar to any other paper I've been writing. You make an argument--Professional statement, annotations--and you support your argument--artifacts"). Some students linked the process of creating a portfolio with having one, by describing the portfolio they created as a starting point for developing future portfolios such as in the following quote: "I am taking away a foundation for a future engineering portfolio if I wanted to start one, as this one is a solid foundation then only polishing would be needed at this point." It was interesting to note the number of portfolio-only responses associated with the CBP condition.

## Discussion and Concluding remarks

In this paper, we showed how the three dimensions of disciplinary knowledge, identity, and navigation can be used to make sense of what students reported taking away from two types of portfolio construction activities. We were able to code much of the data using these dimensions and then identify themes within each dimension. For example, students reported greater awareness of their own knowledge and how to think about what is accountable at any point, stronger identification with engineering, and a broadened sense of what could be included in the engineering pathway to date and ideas about how to manage the pathway in the future. These results suggest that portfolios can help students take greater responsibility for their education. The relatively small scale of these interventions (particularly when compared to large scale curricular change) makes these results even more compelling. The comparative emphasis of this study is intended to provide information for creating portfolio assignments that meet available resource requirements (i.e., the amount of time available) and also help students focus on important learning goals. In our work, the results suggest that educators who are interested in helping students think directly about identity issues should consider the cross-curricular portfolio over the course-based portfolio.

We believe the results presented here may underestimate the extent to which students had insights relative to each dimension since we analyzed only a small amount of information from each student. Because our questionnaire was more extensive, we will be able to explore this conjecture in future analyses. What we will not be able to explore directly is whether students would have had these insights without the portfolio intervention. While it does seem plausible that some students would have such insights eventually, it nevertheless seems valuable to know that portfolio construction can help students along in this process. This work contributes to the scholarship on portfolios as an educational intervention, efforts to understand engineering learning broadly, and, of particular importance, efforts to help empower engineering students to take charge of their education.

## References

Barrett, H. (2007), "Researching Electronic Portfolios and Learner Engagement: The REFLECT Initiative," Journal of Adolescent & Adult Literacy, 50(6), 436–449.

Campbell, M. and Schmidt, K. (2005), "Polaris: An Undergraduate Online Portfolio System that Encourages Personal Reflection and Career Planning," International Journal of Engineering Education, 21(5), 931-942.

Eliot, M. and Turns, J. (accepted). Constructing professional portfolios: Sense-making and professional identity development for engineering undergraduates. Submitted to the Journal of Engineering Education.

Eris, O. (2007), "Insisting on Truth at the Expense of Conceptualization: Can Engineering Portfolios Help?," International Journal of Engineering Education, 22 (3), 551-559.

Guan, Z. and Turns, J. (2007). Same intervention, different effect: A comparison of the impact of portfolio creation on students' professional development, Proceedings of the American Society of Engineering Education Conference, Honolulu, HI, June.

Knott, E.W., W.K. Lohani, O.H. Griffin, G.V. Loganathan, G.T. Ade, and T.M. Wildman (2004), "Bridges for Engineering Education: Exploring ePortfolios in Engineering Education at Virginia Tech," Proceedings of the American Society for Engineering Education Annual Meeting.

Lappenbusch, S. and Turns, J. (2005). "Finding Their Place in TC: Using a Community of Practice Model to Research Emerging TC Professionals," Proceedings of the International Professional Communicator's Conference, Limerick, Ireland, 524-533.

Lave, J. and Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge University Press: New York.

Stevens, R., O'Connor, K., Garrison, L., Jocuns, A., and Amos, D. (2008). Becoming an Engineer: Toward a Three Dimensional View of Engineering Learning, Journal of Engineering Education, 97(3), 355-368.

Turns, J., Cuddihy, E., and Guan, Z. (submitted). I thought this was going to be a waste of time: Using portfolio construction to support reflection on project-based experiences, Submitted to the Interdisciplinary Journal of Problem-based Learning.

Williams, J.M. (2002), "The Engineering Portfolio: Communication, Reflection, and Student Learning Outcomes Assessment," International Journal of Engineering Education, 18(2), 199-207.

## Acknowledgments

This work has been supported by the National Science Foundation through grant REC-0238392. The authors wish to thank the students who participated in this study and shared their experiences with us. The authors wish to thank Matt Eliot and Cindy Atman for their contributions to this work.

# Spatial and Temporal Embedding for Science Inquiry: An Empirical Study of Student Learning

Tom Moher, Jennifer Wiley, Allison Jaeger, Brenda López Silva, Francesco Novellis
University of Illinois at Chicago, 851 S. Morgan (M/C 152), Chicago, IL 60607, USA
Email: moher@uic.edu, jwiley@uic.edu, ajaege1@uic.edu, brendita@uic.edu, fnovel2@uic.edu
Deborah Kilb, Scripps Institution of Oceanography, IGPP/UCSD, La Jolla, CA 92093, USA
Email: dkilb@epicenter.ucsd.edu

**Abstract:** In consecutive years, a fifth grade teacher of a self-contained classroom enacted five-week Earth sciences units that included learning activities focusing on the interpretation of seismograms and the location of earthquake epicenters. In one class, the unit utilized an *embedded* design that situated learners within the spatial and temporal extent of the phenomenon during the epicenter determination activities. In the other class, while the activity set remained the same, the embedding features were removed. Students in the embedded condition demonstrated greater learning gains than their non-embedded counterparts in pre-post assessments of student skill, declarative knowledge, and conceptual understandings, even among topics unrelated to the determination of epicenters. Post-activity student interviews evidenced strong preference for the immersive, asynchronous, and temporal staging components of the embedded condition.

## Introduction

In designing classroom-based learning activities to support science inquiry, one of the most significant challenges is the choice of strategies through which learners might engage with the phenomenon under study. While technologies as old as the printed page have expanded the modalities of engagement and broadened the range of accessible phenomena, they sometimes do so at the cost of some distancing of the learner from the phenomena themselves. As a consequence, learners sometimes lose the opportunity to experience some of the challenges—and excitement—inherent in conducting investigations.

In this paper we report the results of a quasi-experimental empirical study of learner outcomes associated with an activity structure that seeks to shape learner inquiry experiences by *embedding* them within the spatial and temporal extents of the scientific phenomenon of a series of earthquakes. Spatial embedding is implemented by introducing the conceit that the classroom itself is a (scaled) area of seismological activity situated on a tectonic plate boundary. Over the course of about a month, learners experience a sequence of simulated earthquakes—"RoomQuakes"—at unpredicted times, embedding them within the (scaled) temporal course of the phenomena. Students are challenged to locate the plate boundary bisecting the classroom by reading simulated seismograms from seismometers situated around the classroom and using trilateration to combine those results in a way that permit them to determine the epicenters of the seismological events.

In previous reports, we have described the "embedded phenomena" framework (Moher, 2006) and described learner outcomes for a similar RoomQuake unit relative to a no-treatment control (Moher, 2008). Here we extend that work by contrasting the learner outcomes of a fifth grade classroom engaged in an Earth sciences unit on earthquakes incorporating the embedding strategies described above with those of a second class, taught by the same teacher, that received the same instruction and enacts the same set of activities within a *non-embedded* framework. In this second condition, the seismological inquiry is centered on the determination of a fault line in southern California based on the sequential massed analysis of an historical series of earthquakes, using maps rather than the classroom space as the locus of trilateration. Learners were assessed before and after the unit across multiple outcome spaces, including skill in the interpretation of seismograms the use of trilateration, and understandings of the temporal, spatial, and intensity distributions of earthquakes. We also assessed structural and causal content domain understandings associated with the surrounding instructional unit, such as the structure of the Earth and the causes of earthquakes, that were not directly associated with the fault-line determination activities. After the unit, we probed the impact of embedding by posing hypothetical choices between the treatment learners had received and the alternative treatment. In the following we report on these outcomes and discuss their implications for our understandings of how specific elements of the "embedding" might impact learning.

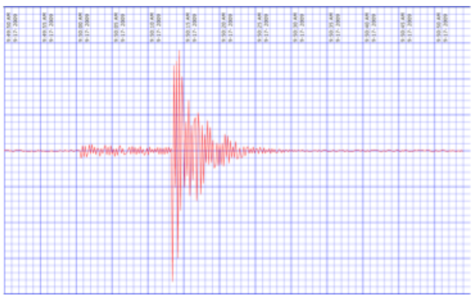


Figure 1. Simulated strip-chart seismogram used in RoomQuake.

## RoomQuake

In the embedded version of RoomQuake, students adopt the pretense that their classroom is an active seismic field, and that a series of earthquakes is expected over the course of several weeks within that field. Computers running conventional browsers, served from our laboratory, act as simulated seismographs (Figure 1) that depict continuous strip-chart recordings (seismograms) of local vibration, where locality is conditioned upon their specific placement of the computers within the classroom. Most of the time, the seismograms reflect a low level of background vibration. At (apparently) unpredictable times, rumbling speakers at each seismograph signal the onset of an earthquake. Upon this signal (or as soon thereafter as classroom instruction permits), students move to the seismographic stations to read the waveforms.

Reading the seismogram recorded at a single location provides two critical pieces of information: the magnitude of the event, and the distance (but not direction) of the event from the recording station. Determining the epicenter of an earthquake requires readings from multiple sites, which may be combined together through the process of trilateration to obtain a location solution. In the embedded version of RoomQuake, we use calibrated dry-lines anchored at the seismographs to sweep out arcs of potential epicenter loci; the solution is obtained when the students at the end of those lines converge at a common point. Once the location and magnitude have been determined, the teacher hangs a color-coded (representing magnitude) Styrofoam ball from the ceiling at the epicenter point, providing an historical record of the event series, and students update poster-based representations of the temporal and intensity distributions of the events (Figure 2).

Figure 2. Photo of classroom hosting embedded version of RoomQuake. Styrofoam balls on the ceiling mark event epicenters, color-coded by magnitude. Wall charts are used to track timing and magnitude of event series.

## Embedding as an Activity Design Strategy

### Spatial Embedding

A central component of our framework is the conceit that the phenomenon under investigation is unfolding within the confines of the classroom itself. This is not an obvious choice, nor necessarily the correct one. When studying seismology, for example, many learning researchers would argue that authenticity demands that the discourse be situated in a scientifically accurate geographic framing; we should be talking about earthquakes around the Ring of Fire, for example. We conjecture, however, that by situating the phenomena "in here" rather than "out there" we might increase learners' emotional interest in the phenomena, and leverage incidental associations between the simulated and real worlds (e.g., "the epicenter was right over my desk"). Moreover, we hope that by situating the imaginary phenomena within the classroom space students can build on their accumulated knowledge of the physical, social, and cultural features of the environment as they undertake a new type of activity.

Another feature of our approach is the decision to maximize the nominal spatial extent of imagined phenomena by scaling them (up or down) to fill the physical space of the room. From a perceptual perspective, we hope to increase the salience of the phenomena for learners (Collins et al., 1991; Yantis & Egeth, 1999). On a more practical level, we also believe that this strategy can reduce congestion in the classroom by allowing students to use the entire floor space as they conduct their investigations.

In embedded phenomena, access to the representation of the state of phenomena is physically distributed throughout the space of the classroom. We believe that this offers three important benefits. First, it creates multiple natural contexts for students to engage in discourse with peers and teachers (Vygotsky, 1978) concerning the phenomenon. Second, it reinforces the important science concept that understanding the state of a phenomenon might not be possible from a single observation, but may require multiple probes from different vantage points that require aggregation and coordination to come to full understanding. Third, we expect that by requiring physical movement from one part of the room to another in order to obtain complementary data we might reinforce memory by associating it with a physical action (Wisneski, et al., 1998).

Our strongest motivation, however, draws from a desire to physically immerse learners within the experience (Dede et al., 1997); in our framework, not only are the phenomena embedded in the space, indeed the learners themselves are embedded within the phenomena. The importance of embodiment has strong advocates both in psychology (Johnson, 1987; Clancey, 1997; Clark, 1997; Glenberg, 1997, 1999; Winn, 2003) and human-computer interaction (Dourish, 2001). Embodiment approaches argue that "thought grows from action and that activity is the engine of change" (Thelen, 1995). In this perspective, cognition arises specifically through bodily interactions with the world. Learning that pairs action and knowledge, or engagement in goal-directed actions, particularly within social contexts (Lindblom & Ziemke, 2007), is viewed as necessary for higher cognitive capacities of thought and understanding to develop.

### Temporal Embedding

The embedded phenomena framework engages the issue of time along two important dimensions: duration and persistence. The long time course of these deployments offers three important benefits. First, it opens the door to the study of phenomena that unfold slowly, requiring investigative processes unlike those used in most classroom science work. A second potential benefit lies in the value of time for students to become meaningfully involved in the enterprise of scientific investigation: different learners engage in activities at different paces. Our prior classroom experience (Moher, 2008) led us to expect that while highly motivated, achievement-oriented students would readily become engaged in our activities, other students would need time to move, in Lave and Wenger's terms, from the periphery to the center of the community of scientific practice (Lave & Wenger, 1991). The persistent representation of phenomena, combined with the spatial immersion, further promotes the goal of engaging all students; for all but the most dedicated non-participant, it eventually becomes easier to participate in an activity that impinges on his or her perceptual system all the time, wherever they look, than to ignore it, particularly when respected peers are engaged in the activity. Finally, by spreading interaction with the phenomena over multiple episodes we hope to take advantage of the potential of temporally distributed instruction over “massed” instruction with respect to recall and motor skill development (Donovan & Radosevich, 1999; Cepeda et al., 2006). It is important to note that effects of spacing and learning with respect to more complex material are less well understood (Dempster, 1989).

Persistence brings at least three additional benefits. First, by continually representing phenomena, we create the opportunity to reinforce the concept that, in nature, important state changes are not always synchronized to fixed schedules, that “things happen when they happen” (asynchronously with respect to the flow of instruction), and that scientists (particularly in observational sciences) are often at the mercy of events rather than the other way around. Second, persistence provides opportunities for "incidental" learning in much the same way that foreign vocabulary words adorning classroom walls may result in learning without explicit reference during formal instruction. We argue that the role of "student in the classroom" inherently demands the ability to attend to multiple concurrent threads of activity; at the same time that a teacher is speaking, a student might be working on a laboratory project, avoiding a spit-wad propelled in their direction, tracking the progress of a playground basketball game visible through the classroom window, and receiving an oral invitation to an after-school event from the student at the next desk. By adding an attentional channel that promotes curricular goals, we address the human need for variety and offer a potentially productive alternative to the normative instructional flow. Finally, by interleaving salient simulations with regular instruction, we surreptitiously expand students' opportunities to engage with additional science content.

## Method

We investigated the alternative learning conditions with two cohorts of fifth grade students across subsequent years with the same teacher in a Midwestern U.S. elementary school. The teacher had previously conducted three RoomQuake units. The two cohorts were randomly assigned to condition with 27 students in the embedded condition and 18 students in the non-embedded condition. There was attrition of one student in each condition due to incomplete data.

Both classes received the same lessons during their regular science periods on Earth science concepts including the background on the Earth’s layers and composition, the existence of tectonic plates, and the geological features that relate to the interactions of these plates.

Both classes also completed a series of 15 earthquake activities where they computed the epicenter of earthquake events from simulated seismograms. The main manipulation was in how the students engaged in these activities across the two cohorts. Students in the “classroom embedded, temporally-distributed” condition experienced a series of 15 simulated earthquakes (“RoomQuakes”), each presumed to occur within the physical space of the classroom. The simulation was effected by placing four 24” iMac computers around the classroom; each computer served as a simulated seismograph which depicted a continuously running strip chart recorder of ground vibration *for that exact location in the classroom*. Between events, the seismographs displayed simulated random noise; when a RoomQuake occurred, however, they traced out unique characteristic waveform (seismogram) corresponding to the expected vibration at their specific locations due to an event at a particular location in the classroom. The earthquake events occurred at unexpected times during the school day over a four-week period and were signaled by a low-frequency rumbling sound generated by the speakers in the computers.

In the non-embedded condition, students also worked with a series of 15 earthquakes, however, these were not simulated, nor were they presumed to be occurring in the physical space of the classroom. Instead students used the computers to access historical seismogram data from 15 earthquakes that occurred in Southern

California. For this condition, each computer represented a different seismograph station in Southern California and the four iMac computers were placed all in a row. The screens displayed snapshots of single seismogram readings and the students were able to go forward or backward through 15 snapshots representing the 15 different earthquake recordings. Students completed this set of activities as a single unit during their normal science time (it generally took a couple class periods to complete the set).

For each event in both conditions, students working in small teams were responsible for determining the epicenter and (Richter) magnitude of each earthquake and cataloging information about the event. By reading the seismogram at each seismographic station, students were taught how to determine the magnitude of the event, and the distance of the event from that station. They learned that earthquakes generate multiple waves that travel at different rates; and that the latency between arrival of two of those waves, P and S, is proportional to the distance between the earthquake and the seismograph. Magnitude can be determined by comparing event distance with graph amplitude. A reading from a single station, however, is insufficient to locate the epicenter, since it does not provide directional information. In order to obtain a solution, students needed to combine the distance information from multiple (at least three) stations through the process of trilateration.

In professional seismologic practice, trilateration is performed through the process of determining the common point of intersection among three or more circles, requiring a level of mathematical sophistication beyond that held by our learner group. To simplify this process for the students, the embedded class used calibrated dry-lines anchored at each of the seismographic stations to sweep out arcs reflecting the locus of solutions from the individual seismic stations until they found the location in the room where the endpoint of their measures coincided. Alternatively, in the non-embedded condition students used calibrated strings that were anchored at each of the seismographic stations on their large maps of Southern California to find the endpoint where all the measures coincided.

The unit also used workbooks that gave students practice with data recording methods and various kinds of symbolic representations of the data. Three types of public displays were also created. For the embedded group, the locations of the events were recorded by hanging Styrofoam balls from the ceiling at their epicenters, color-coded to reflect the event magnitudes. In the non-embedded group, however, students located the earthquake epicenters on a large map of Southern California and used color-coded stickers to reflect the magnitude. Two other public data representations were also maintained using colored sticky dots on large wall posters: magnitude frequency distribution (the number of earthquakes of magnitude 3, 4, 5 and 6) and a timeline of the events. The embedded condition had access to these representations through out the entire unit, while the non-embedded condition only saw them during science classes when working with the earthquake data.

## Learning Goals and Measures

The learning goals of the RoomQuake unit were focused in three areas: (1) the acquisition of skill in authentic seismological practice, including the determination of event distance and magnitude, and the use of trilateration to determine event epicenters, (2) the development of an understanding of the distributional characteristics of earthquakes across the dimensions of space, intensity, and time, developed through student observation of the patterns reflected in the emerging historical record of events, and (3) an understanding of target Earth science concepts.

*Skill in seismological practice*. Following the unit, students were individually interviewed and asked to demonstrate their ability to read seismograms and compute epicenters. This included locating the arrival of the P and S waves, determining the distance between the arrivals of those waves, and determining the amplitude. They were also asked to find the magnitude of the earthquake using the data they had already collected from the seismogram. Additionally, students were asked to show the loci of potential epicenters using two different methods. First they had to find the epicenter by means of the method used in their class (either with strings in the classroom or strings on a map) and then as a transfer task they were asked to find an epicenter using three compasses on a piece of graph paper. We also probed declarative understanding of seismograms and trilateration through a series of multiple-choice prompts using a written instrument that was administered prior to, and directly following, the instructional unit.

*Understandings of earthquake distribution patterns in space, time and intensity*. Over time, collections of earthquakes exhibit characteristic distributions in where they occur (along fault lines marking plate boundaries), in when they occur (after-shocks follow large seismic events), and in their magnitude (strong earthquakes are less frequent than mild earthquakes). We probed these understandings through a series of multiple-choice prompts using a written instrument that was administered prior to, and directly following, the instructional unit. The same test was administered to both classes.

*Structural and Causal Earth Science Concepts.* Also included in the written instrument administered prior to, and directly following, the instructional unit we assessed understanding of Earth science concepts with a series of multiple-choice questions. The questions probed knowledge on topics including Earth structure and composition, plate tectonics, and the formation and location of geological features at plate boundaries.

## Results

### Pre/post Comparisons

The results of the pre- and post-activity assessments are shown below in Figure 3.

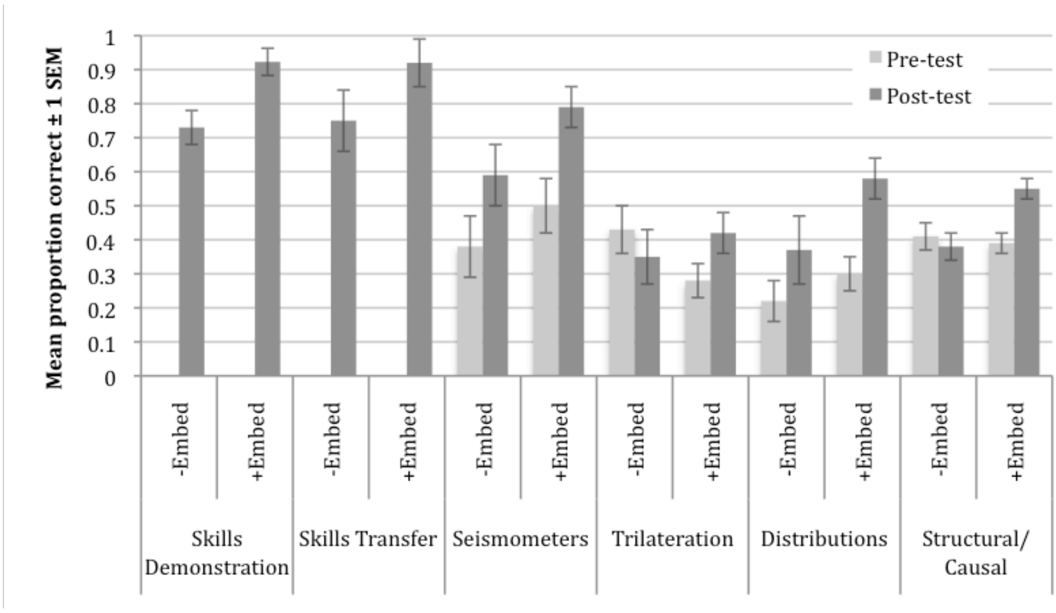



Figure 3. Performance on pre- and post-activity assessments for students in the embedded ("+Embed") and non-embedded ("-Embed") conditions: demonstrations of ability to identify earthquake epicenters, understandings of seismometers and their use, trilateration, earthquake distributions, Earth structures and causes of earthquakes.

*Demonstration of Epicenters.* For this task, students were asked to complete the epicenter tasks using the same methods as they had during RoomQuake (strings in classroom or on maps), as well as to compute epicenters using compasses on graph paper as a transfer task. The results were analyzed with a 2 (embedded, non-embedded) x 2 (same context, transfer context) repeated measures ANOVA. Neither the main effect for context, nor the interaction, were significant. However, the main effect for embeddedness was, $F$ (1, 40) = 5.34, $MSE$ = .125, $p$ < .03. Students in the embedded condition outperformed students in the non-embedded condition on both practiced and transfer skill assessments.

For the remaining assessments, the results are analyzed with 2 (embedded, non-embedded) x 2 (pretest, posttest) repeated measures ANOVAs with pretest and posttest serving as the repeated measures.

*Seismometer Use.* There was a significant main effect for test time, $F$ (1, 42) = 12.7, $MSE$ = .10, $p$ < .01. Students gained from pre- to posttest in their understanding of seismometers and their use. There was a marginal effect for embeddedness, as students in the embedded condition tended to have better understanding of seismometer use, $F$ (1, 42) = 3.76, $MSE$ = .15, $p$ < .06. The interaction, however, was not significant, with similar performance gains across the two conditions.

*Understanding Trilateration.* There was no main effect for embeddedness condition or test time. However, the interaction of these variables did reach significance, $F$ (1, 42) = 3.97, $MSE$ = .06, $p$ < .05 with students in the embedded condition showing greater gains than students in the non-embedded condition.

*Earthquake Distributions.* There was a significant main effect for test time, $F$ (1, 42) = 18.4, $MSE$ = .06, $p$ < .01. Students in both conditions gained from pre- to posttest in their understanding of earthquake distributions. However, neither the main effect for condition, nor the interaction, reached significance.

*Structural and Causal Earth Science Concepts.* Overall there was a significant main effect from pre- to post-test on the understanding of Earth science concepts, $F$ (1, 42) = 5.88, $MSE$ = .02, $p$ < .02, and a marginal effect for embeddedness condition, $F$ (1, 42) = 3.79, $MSE$ = .02, $p$ < .06. These were both qualified by a significant interaction, with students in the embedded condition showing greater performance gains, $F$ (1, 42) = 7.76, $MSE$ = .02, $p$ < .01.

## Student interviews

An examination of the students' responses to the open-ended questions posed during the post-unit interview gives some additional insight into the features of the design that impacted their perception of the activity. In response to an item asking learners to describe "what they did" during the unit, students in the embedded condition focused strongly on the epicenter determination activity, while their non-embedded counterparts were much more likely to respond with descriptions of the other curricular activities included in the unit. To a question about the length of the unit intended to gauge student patience and boredom with the unit, "embedded" students were more likely to express a desire that the unit had been longer, with more "non-embedded" students voicing the opinion that it was too long.

Included in the interviews were several items that asked students to make comparisons between the condition that they experienced and hypothetical alternatives characteristic of the alternative treatment with respect to which would "help you learn about earthquakes and their causes better." In the following we will focus on the responses provided by the students in the embedded condition.

In one item, we asked students whether they would have preferred to have all of the earthquakes "back to back" on consecutive days (the "massed" treatment). In their responses, 22 of the 25 students interviewed indicated a preference for the temporally distributed approach, characterizing it as affording better opportunities for learning.

> Charley: *"The one we did, like which was when you guys did RoomQuakes because we took the time on each one so it helped me learn. I can't learn really fast."*

Asynchrony was perhaps the most appealing design element from the perspective of the learners, all 25 students saying they would learn better from the random earthquakes than ones scheduled at specific times, as suggested in another interview item. In their responses, they focused on realism and the excitement of not knowing when an earthquake was coming:

> Marty: *"It did, because if they were only during science time, that's not how they happen in real life, they only happen in real life, they don't have a certain schedule. In real life, they don't have a certain time that they are going to do it, so that gives you practice for the real world."*
>
> Robert: *"Being a mystery and at different times because, well like having a mystery, because for when there is a real earthquake coming you don't even know so you're just doing your work and it's always a mystery."*
>
> Danielle: *"Because of you did them all in a row in a couple days then we would know what to expect and when it would happen, but when you guys did it, it threw us off guard completely."*
>
> Charley: *"It just catches you in headlights, like you really don't know when it is going to come. It just surprises you."*

When asked whether the activity might have been more valuable as a learning activity if the trilateration work were done on a map rather than in the physical space of the room, 19 students preferred the latter, characterizing is as more "hands-on," more accurate, and more fun that a map version.

> Jacob: *"In the room because it would be better because it was turning our room into, say, California and we were having earthquakes."*
>
> Ariel: *"I think putting it on the room because it's, you can locate it more easily. Like it is easier to locate. Because you can like see where it is exactly like maybe you might, sometimes you could be off on a map."*

In their responses to this item, several of the students described the visceral impact of the experience in terms related to the presence construct.

> Abby: *"I like the way we did it here, it is like more real life because there are like RoomQuakes actually happening, it is not like we are reading what happens, we are actually in it."*
>
> Nguyen: *"Right here because I can feel it and I can do it right here."*

Jacob's response anticipated the final question in our interview, which asked students if it would be better to characterize the earthquakes as "happening right here in your classroom" or "somewhere else, like southern California." Over half the students favored the embedded treatment, saying that they could "experience more" and that it would be better for learning.

> Michelle: *"Well, I think always visualizing it, more experiencing it is better for people to learn because doing, being able to do it is not the same as being able to see it or hear it or because you will be able to interact with it and you'll be able to feel it. So, it's like an actual earthquake. But, if it is in Southern California you can't feel it and you really don't know if it happened."*
>
> Shaquille: *"Probably in by our classroom because if it is in another state or something, like, it's like you don't know how it is or how it's like and a RoomQuake is kind of like an earthquake and you can actually feel how it sounds like."*

In contrast, six of the students preferred the hypothetical alternative treatment, characterizing is as "more realistic," a better way to learn more about other places, and more fun to imagine being somewhere far way. The remaining students expressed no preference.

## Discussion

While the number of subjects and variability of performance limited the power of the design, taken as a whole the study provides support for the claim that the spatial and temporal embedding elements employed in RoomQuake had a positive impact on learner outcomes. Students in the embedded condition consistently outscored their non-embedded counterparts on post-test assessments, with significant gains in several areas.

At the outset of the instructional units, none of the students had prior experience with the interpretation of seismograms or the use of trilateration to resolve distributed readings. Given the multiple opportunities for developing these skills, we were not surprised to see that students were proficient at these tasks at the conclusion of the units. We were surprised, though, at the significantly better performance demonstrated by the students in the embedded condition, since students in both treatments were given the same number of practice opportunities. The difference in performance cannot be attributed to a mismatch between the practice during the unit and during the assessment, since students used the same media for demonstrating the trilateration process (either calibrated dry-lines or strings on a map) that they had used in class in learning the skills.

What then, could account for the difference? While the classroom teacher organized students into teams for the epicenter determination activities and encouraged students to rotate roles within those groups, ultimately the choice of whether and how to participate rested with the individual learners. One possible explanation is that the opportunity to engage in the activity might have been diminished in the non-embedded case because it was more difficult to share responsibilities; determination of the epicenter, for example, required only a single student in the non-embedded condition whereas a pair of students (one anchored at the seismograph, one sweeping out arcs at the other end of the dry-line) was required in the embedded condition. An alternative explanation is that the motivation to participate in the non-embedded condition was damped by the fatigue or boredom induced by the massed scheduling, predictability, or less physical nature of the activity, or by the more restrained "impact" of the fruits of their labor on the accumulating inquiry evidence base.

The significant pre-post gains on the items related to seismometers and their uses, we believe, might be direct outcomes of the real-time and embodied nature of the embedded condition. In the embedded condition, students had the opportunity to witness the differential arrival of P- and S-waves at the simulated seismometer both visually (a change from "flat-line" to higher-amplitude graphs) and aurally (the rumble accompanying the arrival of the first wave). This may have reinforced their understandings of the seismometers as responsive instruments whose relative performance was dependent on their (differential) distances from the simulated seismic events.

Post-test performance on the trilateration items showed only weak gains for students in the embedded condition, and in fact a small loss for students in the non-embedded condition. We suspect that these results may be an artifact of the assessment items, which situated students in a transfer task requiring them to identify the loci of points defined by cartoon characters looking for one, two, and three items located around them, a task some students had difficulty in interpreting.

Students in both groups improved in their declarative understandings of the distributions of earthquakes, with students in the embedded group showing about twice the level of improvement from pre- to post-test assessment. Here we suspect that the salience of the representations of the empirical data (Styrofoam balls hanging from the ceiling and large posters with representations of the formative data at the front of the room), combined with the longer time course of the data collection process, worked to learners' advantage in recognizing the emerging patterns.

The most surprising result was the significance of the difference in learning gains on the Earth science concepts (structure and causes of earthquakes) between the two groups. It is important to keep in mind that the learning activities associated with these concepts were identical in both classes, and were decoupled from the activity associated with the determination of event epicenters, in that the earthquakes themselves were merely the surface manifestations of underlying structures and processes not explicitly represented by our technologies.

While the "embedded phenomenon" strategy incurs minimal technology and materials costs, its use in classrooms demands significant preparation and scaffolding effort on the part of teachers. A necessary prerequisite to broader use of the technique is the existence of an evidentiary base that associates its use with positive learner outcomes across teachers and science content domains. The research presented here represents the first products of an ongoing counter-balanced multi-classroom, multi-domain research study that will be completed in 2010.

## References


Collins, A., Brown, J., and Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, *6*(11), 38-46.

Cepeda, N J, Pashler, H, Vul, E, Wixted, J, & Rohrer, D (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*.

Clancey, W. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge, MA: Cambridge University Press.

Dede, C., Salzman, M., Loftin, R. B., & Ash, K. (1997). Using virtual reality technology to convey abstract scientific concepts. In Jacobson, M. J., Kozma, R. B. (Ed.), *Learning the Sciences of the 21st Century.* Lawrence Erlbaum.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, 1, 309-330.

Donovan, J., & Radosevich, D. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795-805.

Dourish, P. (2001). *Where The Action Is: The Foundations of Embodied Interaction*, MIT Press.

Glenberg, A. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1-55.

Glenberg, A. (1999). Why Mental Models Must Be Embodied. In *Mental Models in Discourse Processing and Reasoning*, Rickheit, G. and Habel, C. (eds). New York: Elsevier.

Johnson, M. L. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago University Press.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.

Lindblom, J., & Ziemke, T. (2007). Embodiment and Social Interaction: A Cognitive Science Perspective. In: Ziemke, Zlatev & Frank (eds.). *Body, Language and Mind. Vol. 1: Embodiment* (pp. 129-162). Berlin: Mouton de Gruyter.

Moher, T. (2006). Embedded Phenomena: Supporting Science Learning with Classroom-sized Distributed Simulations. *Proceedings ACM Conference on Human Factors in Computing Systems (CHI 2006)* (Montreal, Canada, April 2006), 691-700.

Moher, T. (2008). Learning and participation in a persistent whole-classroom seismology simulation. *Proceedings International Conference of the Learning Sciences* (ICLS 2008), Vol. 2 (Utrecht, Netherlands, June 2008), 82-90.

Thelen, E. (1995). Time-scale dynamics in the development of an embodied cognition. In *Mind In Motion*, R. Port and T. van Gelder, Eds.. Cambridge, MA: MIT Press.

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*(4), 625-636.

Winn, W.D. (2003). Learning in artificial environments: Embodiment, embeddedness and dynamic adaptation. *Technology, Instruction, Cognition and Learning*, 1, 87-114.

Wisneski, C., lshii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., & Yarin, P. (1998). Ambient displays: Turning architectural space into an interface between people and digital information. *Proceedings International Workshop on Cooperative Buildings (CoBuild '98)* (Darmstadt, Germany, February 1998), 22-32.

Yantis, S. & Egeth, H. (1999). On the Distinction Between Visual Salience and Stimulus-Driven Attentional Capture. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 661-676.


## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DRL-0735569. The authors wish to acknowledge the helpful comments and suggestions offered by the anonymous referees and to express our appreciation to our partner learners and teachers for their help and patience.

# Appropriating Conceptual Representations: A Case of Transfer in a Middle School Science Teacher

Suparna Sinha, Steven Gray, Cindy E. Hmelo-Silver, Rebecca Jordan, Sameer Honwad,
Catherine Eberbach, Rutgers University, 10 Seminary Place, New Brunswick, NJ 08901
suparna.sinha@gse.rutgers.edu, stevengray142@yahoo.com, cindy.hmelo-silver@gse.rutgers.edu,
rebeccacjordan@yahoo.com, sameer.honwad@gse.rutgers.edu , catherine.eberbach@gse.rutgers.edu
Spencer Rugaber, Swaroop Vattam, Ashok Goel, Design & Intelligence Laboratory, School of
Interactive Computing, Georgia Institute of Technology,Atlanta, GA 30332
spencer@cc.gatech.edu, svattam@gmail.com, goel@cc.gatech.edu

**Abstract:** This paper presents a case study in which a middle school science teacher modifies her classroom instruction and teaching materials using the Structure, Behavior, and Function (SBF) representational framework to transfer her reasoning about one natural system (an aquarium ecosystem) to another natural system (human cells and body systems).

Promoting an understanding of complex systems is a difficult yet important component of scientific literacy (Sabelli, 2006). Through emergent processes and localized interactions, the behaviors of a system's components affect its overall function (Jacobson & Wilensky, 2006). These interactions are often dynamic and invisible which make them difficult for learners to understand and presents challenges for teachers (Feltovich et al.,2001 Hmelo-Silver et al., 2007).

The Structure-Behavior-Function (SBF) conceptual representation allows students to understand the relationships between the structures, behaviors, and functions of systems and to map SBF across multiple levels. In the science classroom, academic researchers have developed tools utilizing the SBF representation as a way to teach about multiple complex systems (Hmelo-Silver et al., 2007; Liu & Hmelo-Silver, 2009). Structures are defined as the components of a system, behaviors as the mechanisms or processes that occur within a system and functions as system outputs (Goel et al, 1996; 2009). We present a case study of a teacher who appropriated the SBF conceptual representation that was originally introduced by a team of researchers, to create tools to meet her curricular needs. The focus of the study is to understand how the teacher transferred the use of the SBF representation between two different curriculum units. Some of the common critiques against the classical approach of transfer have put the spotlight on questions related to the conditions that facilitate transfer, the context of instruction, and the pertinent cues that learners need to be able to identify that signal the application of appropriate problem solving strategies (Lave, 1988, Bransford et al., 2000; Mestre, 2003). The study looks at transfer under two alternate perspectives and the possible synergy between using multiple perspectives to study transfer.

Our study considers transfer from both an actor oriented approach (Lobato, 2002, 2006) and a preparation for future learning perspective (Bransford & Schwartz, 1999) to investigate the teacher as learner applying knowledge in a new area. We look at the influence of previous activities and procedures adopted by a learner to construe similarities in new situations (Lobato, 2006). In particular our interest is in understanding how a teacher constructs the relationships and similarities between two systems; one provided by researchers and one designated by the teacher. Adopting an actor oriented perspective requires considering which connections the learners make, on what basis, and how and why those connections are productive (Lobato, 2002). Evidence for transfer from this perspective is found by scrutinizing a given activity for any indication of influence from previous activities. Additionally, we investigate how a greater understanding of SBF might have contributed to transfer from a preparation for future learning (PFL) perspective. This provides a framework for evaluating the quality of particular kinds of learning experiences (Bransford & Schwartz, 1999). The PFL perspective allows us to uncover strategies that have been adopted to interpret a new situation; "Interpretation of the situation invariably involves some use of a previous experience it cannot be reduced to a simple replication of that experience" (Broudy, 1977, p. 11). Integration of these two perspectives on transfer was crucial to understanding the importance of activities and experiences (from an actor oriented perspective) which prepare the learner to develop new concepts that, in this case, focuses on transfer between complex systems.

## A Case of Transfer: The Instructional Context

The context for this study is a technology-intensive curriculum unit that helps middle school science students learn about an aquarium ecosystem using SBF as a conceptual representation. The learning environment is comprised of two parts; (1) Reptools toolkit and (2) The Aquarium Construction Toolkit (ACT). The RepTools toolkit includes a function-oriented hypermedia (Liu et al, 2007) and NetLogo computer simulations (Wilensky & Reisman, 2006). The hypermedia (Figure 1) introduces the aquarium system with a focus on function and provides linkages between structural, behavioral and functional levels of aquariums. In addition, two NetLogo simulations are included to make the macro and micro level behaviors visible to the learners. The second component to the learning environment, ACT, allows students an opportunity to organize their understanding about complex systems by explicitly identifying structures, behaviors, and functions in both an SBF table (Figure 3) and through concept mapping which allows relationships to become explicit (Figure 4). This gives the students an opportunity to understand both individual mechanisms and the meta-level concepts related to complex systems (Goel et al, 2009)

In this study, we highlight a particular case in which a teacher participant, Ms. Y, who has been working with the SBF Aquarium curriculum for more than five years, developed her own instructional tools using the SBF representation and computer-tools to meet her classroom teaching needs. In the fourth year of enacting the SBF Aquarium curriculum, Ms Y used the SBF representation as an instructional tool to teach about the cell and human body systems (digestive, reproductive, respiratory etc.). The result, in collaboration with another science teacher from the same school, Ms. T, was the creation of a human body system hypermedia modeled after the function-centered aquatic hypermedia.

## Methods

To characterize how the SBF representation is appropriated by Ms. Y and to understand the processes by which the teacher transfers these tools to other domains, we adopt a qualitative approach. An interview with the teacher and video analysis sessions were conducted to understand (1) why the teacher transferred her understanding of SBF to new instructional domains and (2) how she transferred these understandings. We held an interview session with Ms Y after she had completed teaching about both the systems. The primary focus of the interview was to understand how she conceived the idea of extending the computer-based representational tools beyond what was expected from her, the influence of her prior knowledge during this process and her attempts to prepare herself to solve new challenges. After the interview transcription, we looked for evidence of mechanisms for transfer and how the teacher constructed similarities between aquarium and body systems.

To understand the nature of transfer in Ms. Y's teaching of the content, we selected representative clips of critical events from Ms. Y's classroom that demonstrates evidence for using the SBF as a tool to teach about another complex system. These video clips included whole class discussions that Ms. Y has with her students while (1) introducing the SBF concept for the aquatic ecosystem the year before she created the cell/body unit (2) introducing the SBF concept for the aquatic ecosystem the year she employed her cell/body unit and (3) explanation of SBF and modeling of the cell/body unit. These clips include three classroom interactions for all three time interactions under evaluation. These video data were analyzed using Interaction Analysis (IA), which involves collaborative viewing of video clips of a group of researchers to examine the details of social interaction (Jordan & Henderson, 1995). The basic goal of IA methodology is to use the video data to understand the means by which people interact in social environments and how, if any, learning takes place. We conducted nine IA sessions successively to collaboratively review the selected video clips, describe observations, and generate hypotheses.

## Results

### Teacher Created Hypermedia

Ms Y., in collaboration with her colleague Ms. T, created a hypermedia in the form of an interactive PowerPoint of the cell and body systems mirroring the aquarium hypermedia developed by the research team (Figures 1 and 2). The teacher hypermedia outlines the different structures in the system along with leading 'why' and 'how' questions. The 'how' questions are directed towards understanding the behavioral aspect of system components and the 'why' questions focus on functions. The teacher created hypermedia was developed as additional learning resources to textbooks to connect cell systems to higher order body systems. Neither the body system hypermedia nor the use of modeling these systems using the ACT software was planned by the research team; the teachers did this of their own volition.

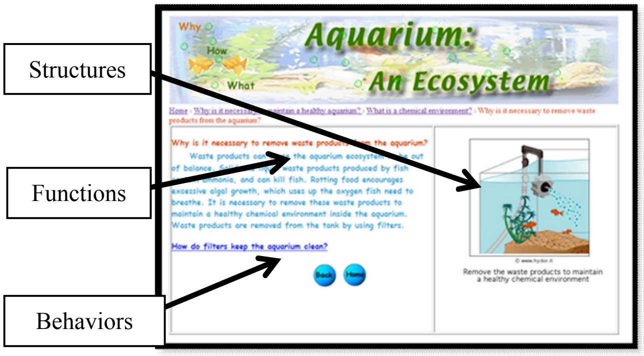


Figure 1. Researcher-developed hypermedia

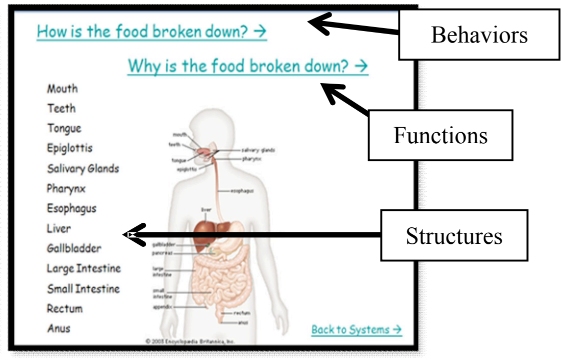


Figure 2. Teacher-developed Hypermedia

## Identifying similarities through the SBF lens

Ms Y's initiative to extend and appropriate our research and develop additional classroom tools led us to consider the process of transfer. We intended to understand the processes she adopted to draw similarities between what she had been teaching for several years (the aquatic ecosystem) to the current unit she developed (cell and body systems). Adopting an actor oriented perspective helped us to articulate which connections she makes, on what basis, and how and why those connections are productive (Lobato, 2002). A PFL approach allowed us to understand how her learning experiences may have prepared her to apply the SBF representation when she created her new hypermedia. Our hypothesis, based on the interview, is that she used the SBF as a tool to trace similarities across complex systems. For example, consider the response when both Ms Y and Ms T were asked about the utility of their hypermedia:

**Researcher**:…what is the purpose of the Hypermedia you guys came up with? And how does it help you teach about complex systems?

**Ms. Y.**: Okay, first we were teaching the cell. And we also have to teach body systems especially in the new standards, so its going to work out that we're actually covering both of those. But they teach the cell as an isolated entity. And we were saying that you know the whole body is made of cells and you know you teach the systems separate, you teach the cell separate, and then we just name the parts of the cell and tell what each part does. And there's no connection about how the parts interact, interact with each other in the cell, let alone how the cell interacts with the whole body.

**Ms. T**: Because kids don't even know that each- Like you're bodies is made up of cells, they just think of it as cells.

**Researcher**: Everything discrete?

**Ms. T**: Right. .., everything's separate and that you know, the idea that everything is made up of different types of cells and the cells work together; they just don't get that, at all.

**Ms. Y**: Right, and it's a hard concept to get. So, what we were thinking about is like the kids actually think when they eat food it breaks down and then leaves the body. They don't get that the food has to go to the cells and the cell actually works and creates energy from this food and then there's a waste and it sends that back to the body for it to be excreted. So we're trying to give them not only the names of the parts and what each part does individually but how it needs to work-

**Ms. T**: I would say we're trying to bring in the behavior into it. They get the structure, we teach part the function, but there was never that behavior why does it need to do what it does….

**Ms. Y**: And we're doing the behavior not only of the cell itself but behavior of all the systems and then the behavior of the whole body. And the cells are all part of that whole body.

This excerpt highlights that Ms. Y understood that the cells are an integral part of the body systems and cannot be taught in isolation. She discusses that systems in the body are not disconnected and have complex mechanisms which allows for higher order operation. It is apparent that she understands how structures within a system perform multiple behaviors in order for it to function effectively. Owing to the strong connection to the SBF representation, we felt it critical to observe the learning trajectory she adopts to refine her understanding of the SBF as a conceptual tool. This led us to analyze Ms. Y introducing the SBF concept (via IA sessions) over multiple years to determine how she may have refined her own understanding of SBF as a basis to then extend it into a new area.

## Refining SBF as a Conceptual Tool

Upon examining Ms. Y's classroom introduction of SBF over time, analysis of the video suggested that her understanding of the SBF representation as a conceptual tool changed. She adopted several distinct strategies to introduce the topic of complex systems ranging from discrete (the year before she developed her unit), to acknowledging complexity (the year she developed her unit), and finally providing a systems perspective with her new cell/body unit. The year before developing her new unit, she began her introduction to the SBF representation by mentioning the new terminology being used to discuss the system, however, she introduced structures, behaviors, and functions as individual constructs that should be dealt with discretely. In the following year, she promoted the SBF representation as a comprehensive perspective, acknowledging it as a conceptual unit. Finally while introducing it within the context of the unit on cells and body she explains it as a system, complete with nested subsystems and stressed the importance of inter-relatedness.

Ms. Y's early introduction to SBF suggests a focus on linear connections. This is shown by the way in which she fills out the ACT SBF table (Figure 4) in front of the classroom. As a way to connect ideas about the Structure, Behaviors, and Functions, she drew clear conceptual lines between one structure at a time and all the behaviors exhibited by that structure as the following example shows:

**Ms. Y:** We just named them all yesterday ... The heater, the fish, the plants. Those things are called the structure. The next word we're gonna use is Behavior. The behavior is what the fish do. What do the things do in the tank? And the next word we're gonna use is Function, okay? So what I want to do today … to start with structure and behavior….. So, I made a chart and …The first column is the structure, or the parts. So everyone write down one of the things in the fish tank is fish. and the second column I wrote was behavior, and the third column I wrote was function. We're going to start with this second column which is behavior. When I ask you the behavior of something, I want to know is what does it do?

**Ms. Y:** "What do fish do?" Swims, eats, breathes, and poops. Okay, all fish swim. That is their behavior okay. They swim. What else do fish do?

**Andy:** Breathe

**Ms. Y:** They breathe. What else do fish do?

**Andy:** Eat..

This excerpt indicates that, after covering structures the day prior, she begins by describing the meaning of the term 'behavior'. This was immediately followed by establishing all linear connections between the structures (fish) and the multiple behaviors (swims, eats, breathes, poops) that this structure exhibits. After promoting an understanding of the behavior exhibited by the structure (fish), she then drew another relationship between each individual behavior to the last column to indicate the behavior's function.

Over time, Ms Y's introduction to the SBF concept became richer and more complex. In the excerpt below taken from the year after she developer her unit, she describes SBF as interconnected entities within a system, rather than discrete elements on a worksheet:

**Ms. Y**: Okay, now, let's do the filter.…. I'm gonna do the filter with you and then you're gonna do one on your own. Alright, so… what does the filter do? What does the filter do? Jim what does the filter do?

**Jim**: Um, cleans out the tank

**Ms. Y**: Cleans the tank. Or cleans the "what part of the tank?"

**Jim**: The water in the tank?

**Ms. Y**: All right, so the filter will clean the water. Okay? Now, why does it clean the water?

**Jim:** So it can put more oxygen into the water?

**Ms. Y:** No. That's another thing that it does. It actually, because it's spinning around, because it's spinning like this, it's actually, one of the things it does- we could make a second bullet- is it adds oxygen to the water. Now, this part here, why does it do it? First of all, I want to stop right here. The filter is this big grey thing here. Right? Now, first of all, how does it work? What's this big tube doing? *[Points to picture of filter on the screen.]*

**Pat**: Sucking up the water

**Ms. Y:** Sucking up the water. Then the water comes up here, right? And it gets sucked up and it goes back here and it pours back down. When it flushes back over that's when the oxygen from the air can get pulled back into the water. Okay, so how- you said it cleans the water- how does it do this?

**Pat:** Well, it has in the filter. The filter has like chemicals and stuff… well like, (inaudible) right here, we're gonna start with the fish tank. This bag, because right here..

**Ms. Y**: What do you think is in this bag?

**Pat:** Bad stuff

**Ms. Y**: Well, eventually the bad stuff is going to get in here, but actually there's charcoal in here, gravel in here. And then when the water flows through it, can it catch all the big chunks?...Maybe the fish feces and stuff like that? So, and then see how it spins back down here? Water splashes and it's pulling in the oxygen. So now… Alright, so now, why does it clean the water? What is the point of cleaning the water? Because of the what? Right, it keeps clean water for the who?

The class went on to discuss the fish and the plants, and how the filter aerates the tank and what the filter does for the whole system. In this example, when Ms. Y discussed the behaviors (cleans water in tank) and function of the filter (by collecting feces from fish) she was guiding students' answers to S, B, and F simultaneously and filling in the chart appropriately rather than focusing on any one aspect in isolation. Ms. Y also used student answers to generate more questions that linked what (Structure) and why (Function) questions throughout her classroom discussion, highlighting the complexity of how the system works.

Later in the same year, when introducing her unit on the cells, Ms Y emphasized how the SBF within a system work as a whole across multiple levels, not directly, but implicitly by leading questions as seen from the example below:

**Ms. Y:** Eventually what we want [the researchers] to do for us is allow us to model systems within systems. What happens if I can click on the cell and zoom in on that and put the cell parts in there? Because they don't have the ability to zoom right in on that one part, are there any ideas on how to connect the cell through modeling to the other body systems? Because you also want to go and look at the function. What do you think?

**Lucia:** Umm, what about if you like umm put a picture of the cell.

**Ms Y:** Yeah but I want to drive everything to the cell because that's, you know, the whole body operates to get things to the cell, you know that right? But then I also want to show what the cell does inside once you send the food there. So how can I show that part? How can I show it on this graph? Okay. You know how this is a system. The body parts and the cell is its own little mini system, how can I show the stuff inside the cell? Should I circle all the mitochondria right around the cell? Or should I pull the cell out and make that part separate? **…**Okay so here on this side we'll do all the body parts connected to the cell, and over on this corner do you want to do the cell again and show the inside of the cell..

These excerpts demonstrate the process of refining of Ms. Y's thinking about the SBF concept as a conceptual tool. Whereas earlier, her focus was primarily in working with living organisms in the aquarium (fish), she later introduced a new level of complexity by trying to relate the influence of non-living structures on other living organisms. She was still focusing largely on structures but she does make connections to behaviors and functions. In addition, she helped students understand that one structure may have multiple behaviors and functions. Comparing her SBF introduction of the cell system here to that of the aquatic ecosystem in the earlier unit, she presented it to the class as an entire system and not as fragmented pieces as S, B & F with linear connections. In addition, when applying SBF to the cell, Ms. Y introduced a meta-perspective by explicitly explaining that the task was to represent their ideas through modeling. Moving away from the isolated task provided in earlier in the context (i.e., filling out the table by first listing structure followed by behavior, and then function) Ms Y explained that the students are going organize their knowledge in the table graph. By placing emphasis on the modeling tool and providing students with the starting point of the structure, the cell, Ms Y explains that the task is to develop a representation of their ideas about the human body system, using the table to organize their ideas and providing the students with leading questions that she had provided earlier when talking about the SBF in the aquarium unit.

This process of transition led us to believe that Ms Y. was an active learner herself. As she explored the aquarium hypermedia and talked about SBF representations in her class, she frequently asked questions to the research team and her colleague Ms. T., to refine her understanding. We think this experience had two effects. First, it helped her identify and address the gaps in her understanding, which prepared her for future learning. Secondly, it shed light on the processes that she as an actor (learner) use to construct similarities between the aquatic ecosystem and cell system (Lobato, 2002).

## Recognition of Teacher as a Learner

In the interview, Ms. Y indicated that since the beginning of her involvement in the project, her knowledge has continually developed. She explained that she was the primary source by which information was passed from the research team to the students and that, over time, she felt that she has become more competent in this role. Effective learners learn to look critically at their current knowledge and beliefs (Bransford, Brown, & Cocking, 2000) instead of making old responses by simply associating new information to the existing concepts or schemas. In the following excerpts from the interview, she acknowledges her lack of mastery over the content and is aware that she refined her ideas of the SBF representation and the aquarium unit which lead to adoption of the new unit:

> Okay, my knowledge of this still develops every year because it's knowledge that [Project PI] had and it- you know- was her angle on something and then I had to try to understand what was going on in her head. So it's taken me many years of practice and talking to [Project PI], talking to [researchers in the room], to kind of get this. And I still do not feel like I'm really solid on it, but I get it more and more each year.

These statements demonstrate Ms Y as a learner in her classroom as she was looking critically at her current knowledge and beliefs (Novick, 1988). This experience prepares her to deepen her understanding of the content, and revise her ideas as she gathers new information.

## Appropriating Salient Features of the Aquarium Hypermedia

When asked about what parts of the Hypermedia she found useful in her own development, Ms. Y felt that working with the same piece of software for 5 years allowed her to incorporate key features within the hypermedia she created. Although her hypermedia does not possess the technological and conceptual sophistication of the aquarium hypermedia, it prepares her for refining her model along a trajectory of increasing expertise (Figures 3 and 4). Moore & Schwartz (1998) suggest that even if students generate representations that are not entirely correct or are faulty, these experiences help them in noticing critical features of experts' representations. This process is important from an actor-oriented transfer framework, as it enables her to see the connections between two situations by identifying the salient features of one learning environment (Lobato, 2002). The excerpts below demonstrate how she is expanding upon her knowledge to integrate important characteristics from the aquarium hypermedia:

> I would say that I definitely liked how each question lead to another question because that's how we modeled ours was every question gave an answer but then lead to another question and another question and another question.

The above excerpt highlights a key feature of the aquatic hypermedia of linking different concepts via leading questions. The students can learn about the behaviors and functions of different structures by browsing these questions.

Another important feature of the hypermedia is its presentation. Each section comprises of a brief description accompanied by a visual image. The following excerpt highlights the fact that Ms Y used it as a parameter while constructing her hypermedia:

> We also used just short pieces of information because I think the kids get bored if you put too much it's overwhelming. We used pictures and then we also had it not only lead to different the next one and the next one but it bounced back sometimes a design in the hypermedia too.

It is clear that as Ms. Y. identified the relevant features of the aquarium hypermedia. This process of experimentation is also helping her clarify her own thinking (Bransford et al., 1990) about the concepts that she is placing within the new hypermedia contexts. It is notable that she transferred other features of the hypermedia structure beyond SBF, including the use of guiding questions as well as the use of short pieces of text accompanied by simple and relevant graphics.

## Appropriating ACT to Model a New System

In addition to appropriating aspects of the Aquarium Hypermedia, Ms. Y also appropriated the ACT program so that students could model body systems in the same fashion as they had for the aquarium system. After students read through the body system hypermedia, they were instructed to develop SBF tables and concept maps using the ACT software (Figures 3 and 4). The following excerpt highlights Ms Y's intentions of evaluating student understanding when placed into the ACT program:

> …at first she [PI] came and she was just testing the kids knowledge and that I was not really involved and then we started, actually we started originally started talking about the cell and the

> body so that was one of the things then she got the idea of the respiratory system because that was an area she worked in and then she, this slowly develops into you know the NetLogo and the Hypermedia then structure, function, behavior, and I think for me it was all just disjointed all the pieces were here and I was just trying to keep up with her. And then…the ACT program helped a lot because it sort of put everything together for me in the end, like okay, here's all the knowledge that the kids have been getting along the way, here is proof that they got it. And for me it was just a slow process of absorbing everything and you know kind of understanding it until I could you know turn key it and then we could turn around and together make another Hypermedia with it

We think that this exemplifies the importance of the ACT software as a capstone to allow for student understanding of the new system to become explicit. This example from the interview, and the classroom task of modeling body systems in ACT, indicates that Ms. Y possessed the confidence to organize the new ideas generated by her hypermedia into SBF terms using the tool and the importance. Additionally it also highlights her ability to appropriate the ACT tool as the final classroom task to evaluate knowledge generated by the hypermedia as a way to organize student ideas about complex systems.

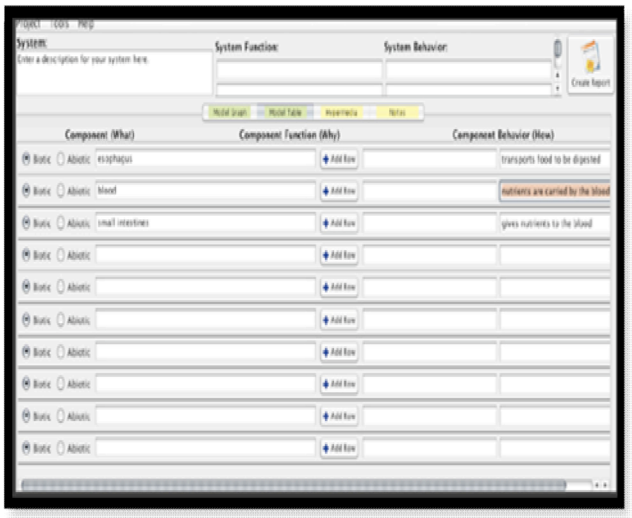



Figure3. Digestive System ACT table view

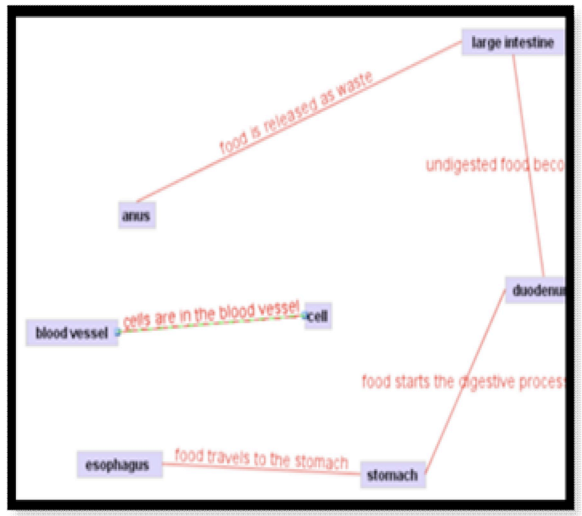



Figure 4 Digestive System Student ACT Model

## Discussion

The most relevant contribution our study makes to the current research on alternative approaches to transfer (Lobato, 2004, 2006; Bransford & Schwartz, 1999) is to further expand on the notion of looking at new ways for understanding learning trajectories. As we observed Ms. Y's transition over multiple years our focus was not in assessing mastery over content knowledge but on the processes she adopted during this transition. In terms of learning trajectories, our results highlight the fact that Ms. Y is looking critically at her current knowledge and gradually evolves to an experienced novice within that content area. Data analysis from earlier years reveals an early understanding of the SBF as a conceptual tool. However, she actively sought resources (fellow colleague, Ms. T; project PI; other researchers present in the classroom) to help her understand the interconnections between SBF. Her increasing confidence in the content area, coupled with collaboration, resulted in her being highly motivated to extend the research tools. From a PFL perspective both these processes are vital as she is able to revise her current knowledge and beliefs and it sets the stage for her to analyze and appreciate critical features of the new information presented to her (Bransford et al., 1990; Moore & Schwartz, 1998). This process of analyzing her own beliefs and strategies also targets the active nature of transfer, which is an important part of PFL. The initiative she took in applying her SBF understanding to teaching a new unit demonstrates her ability to revise and rethink the current situation to suit her current goals. Even the hypermedia she created echoes the same pattern of organization as that of the Aquarium Hypermedia. From a PFL perspective this is valuable as it reveals the importance of activities and practices that are beneficial for "extended learning" rather than on one shot task performances (Bransford & Schwartz, 1999). The analysis presented in this study suggests the possibilities of extended research within the field of nontraditional approaches to transfer. In our case, teacher adoption and appropriation of a learning framework was an exciting by-product of scholarly research because it indicates promising evidence that classroom interventions can be sustained and built upon to be appropriated given specific curricular need. Additionally, our analyses characterize that under the correct conditions, some teachers may be motivated to create their own learning tools when the knowledge is transferred to a new domain, thereby extending the reach of classroom interventions.

## References

Bransford, J. D., Brown, A. L., & Cocking, R. (2000). *How people learn*. Washington DC: National Academy Press.

Bransford, J. D., Vye. N,, Kinzer, C, & Risko, V. (1990). Teaching thinking and content knowledge: Toward an integrated approach. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction: Implications for educational reform* (Vol. 1, pp. 381-413). Hillsdale, NJ: Erlbaum.

Bransford, J. D. & Schwartz, D. L. (1999). Rethinking transfer: Asimple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), Review of research in education (Vol. 24, pp.61–100). Washington, DC: American Educational Research Association.

Broudy, H. S. (1977). Types of knowledge and purposes of education. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 1-17). Hillsdale, NJ: Erlbaum.

Feltovich, P. J., Coulson, R. L., & Spiro, R. J. (2001). Learners' (mis)understanding of important and difficult concepts. InK.D. Forbus&P. J. Feltovich (Eds.), Smartmachines in education: The coming revolution in educational technology (pp. 349–375). Menlo Park, CA: AAAI/MIT Press.

Goel, A., Rugaber, S., & Vattam, S. (2009). Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 23*, 23-55.

Goel, A. K., Gomez de Silva Garza, A., Grué, N., Murdock, J. W., Recker, M. M., & Govinderaj, T. (1996). Towards designing learning environments -I: Exploring how devices work. In C. Fraisson, G. Gauthier & A. Lesgold (Eds.), *Intelligent Tutoring Systems: Lecture notes in computer science*. NY: Springer.

Hmelo-Silver, C. E., Liu, L., Gray, S., Finkelstein, H., & Schwartz, R. (2007). *Enacting things differently: Using NetLogo models to learn about complex systems.* Paper presented at the Biennial meeting of European Association for Research on Learning and Instruction.

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences, 16*, 307-331.

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences, 15*, 11-34.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*, 39-103.

Liu, L., & Hmelo-Silver, C. E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching*.

Lobato, J. (2006). Alternative Perspectives on the Transfer of Learning: History, Issues, and Challenges for Future Research. The Journal of the Learning Sciences, 15, 431-449.

Lobato, J. (2004). Abstraction, situativity, and the "actor-oriented transfer" perspective. In J. Lobato (Chair), Rethinking abstraction and decontextualization in relationship to the "transfer dilemma."Symposium conducted at the annual meeting of the AERA, San Diego, CA.

Moore, J. L., & Schwartz, D. L. (1998). On learning the relationship between quantitative properties and symbolic representations. In *Proceedings of the International Conference of the Learning Sciences* (pp. 209-214). Mahwah, NJ: Erlbaum.

Sabelli, N. (2006). Complexity, technology, science, and education. *Journal of the Learning Sciences, 15*, 5-10.

Schwartz, D. L,, & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475-522.

Wilensky, U. & Reisman, K. (2006). Thinking like a wolf, a sheep or firefly: Learning biology through constructing and testing computational theories – an embodied modeling approach. *Cognition and Instruction, 24*, 171-209.

## Acknowledgments

This research was funded by IES grant # R305A090210 to the third author. Conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of IES.

# Discipline-specific Socialization: A Comparative Study

Iris Tabak, Michael Weinstock, Hilla Zviling-Beiser, Ben Gurion University of the Negev, P.O.B. 653, Beer Sheva, 8415, Israel
itabak@bgu.ac.il, micwein@bgu.ac.il, zvilingh@bgu.ac.il

**Abstract:** Do different disciplines cultivate different epistemologies? We draw on the epistemological framework of D. Kuhn et al. that delineates three perspectives: absolutist, maintaining that knowledge is objective and immutable; multiplist, maintaining a radical relativism; and evaluativist, maintaining a qualified relativism. We conjectured that typical instruction in the humanities would tend to foster evaluativist views more than typical instruction in the sciences. Twenty biology majors and twenty history majors evaluated competing accounts in biology, history and judicial contexts. Structured interviews concerning these accounts were used to assign each participant an epistemological view in each discipline. These results were considered in conjunction with learners' reports of their educational experiences. We were disappointed to find that there was an overall tendency toward absolutism. Our main finding is that students are distinguished by major and epistemological view, and that typical history instruction more than typical science instruction seems to foster evaluativist views in history.

## Introduction

Learners' epistemologies are of increasing concern to learning scientists. What learners believe about the certainty and simplicity of knowledge, and about the source and justification of knowledge influences how they approach knowledge construction and learning (Hofer, 2004; Schauble, Glaser, Duschl, Shulze, & John, 1995; Songer & Linn, 1991). Schooling and level of education appear to be a strong factor in the development of these beliefs (Kuhn, Cheney, & Weinstock, 2000; Schommer, Calvert, Gariglietti, & Bajaj, 1997). Although developmental patterns have been associated with grade level, it seems that educational experiences rather than age explain most differences in epistemological views. Unique educational experiences are associated with epistemological perspectives that depart from that grade-level's norm (Smith, Maclin, Houghton, & Hennessey, 2000), and classrooms with different curricular foci at the same grade level can be characterized by different epistemological belief patterns (Tabak & Weinstock, 2005).

Moreover, there is increasing evidence that epistemological views are discipline-specific, that is, the same individual may hold different beliefs about science than about art or psychology (Hofer, 2006). Despite differences between disciplines, there seems to be some measure of regularity concerning beliefs within and about particular disciplines. This suggests that learning experiences are key in shaping epistemological views, and that different disciplines cultivate different beliefs about the nature of knowledge. In this paper, we report on a study that examined whether different disciplines, college-level biology and college-level history, fostered different approaches to knowledge. Specifically, we compared the ways in which biology majors and history majors evaluated competing-claim scenarios in biology, history and judicial domains. In addition, we explored whether students' approaches to these tasks and any differences among them seem related to students' accounts of their college learning experiences.

## Background

### The Absolutist, Multiplist, Evaluativist Framework

We approach this exploration of biology and history majors' epistemological views from the framework advanced by Kuhn and her colleagues (e.g., Kuhn & Weinstock, 2002). This framework is consistent with a number of approaches in developmental psychology's investigation of personal epistemology (Hofer & Pintrich, 1997). The framework distinguishes between three approaches to knowledge: (1) "absolutist"—the conception of knowledge and knowing as objective and absolute, (2) "multiplist"—regarding all knowledge as subjective and relative and, therefore, indeterminate because of multiple points of view, and (3) "evaluativist"—a qualified subjectivity that integrates subjective and objective aspects of knowledge, permitting a degree of evaluation and judgment of knowledge claims. Table 1 elaborates some of the beliefs about the certainty of knowledge, the source of knowledge and the justification of knowledge that are associated with each of these perspectives.

Table 1: Characteristics of epistemological views along central dimensions of the nature of knowledge

| | Absolutist | Multiplist | Evaluativist |
|---|---|---|---|
| Certainty | Knowledge is certain. Any uncertainty results from error or insufficient information. | There is no certainty. | There is no certainty. But, one account may be more trustworthy if there is more evidence to support it. |
| Simplicity | There is one single correct account based on self-evident facts. | There are multiple accounts. | There are multiple accounts each containing different types of information, arguments, and interpretations. |
| Source | Knowledge comes from experts or authority that are able to discern the objective reality. | Knowledge derives from opinions. | Knowledge is constructed. |
| Justification | Facts either prove or disprove knowledge claims. | Knowledge claims can never be truly defended, because all opinions are equally valid. | Preponderance of valued sources of evidence. |

## Why Adopt this Developmental Framework in the Learning Sciences

This framework has been employed mostly in the area of developmental psychology. In the learning sciences, research on epistemology has focused more closely on particular disciplines, predominantly science, and has focused on the distinction between absolute and dynamic or constructivist views of science/knowledge (e.g., Hogan, 2000; Songer & Linn, 1991). We have found that the Kuhn et al. framework has refined our understanding of epistemological perspectives within particular disciplines, and especially in science. In particular, this conceptualization reveals nuanced distinctions that might exist in students' specific conceptions of the nature of science, by pointing to two perspectives that can fall under a constructivist or dynamic view of science. For example, learners who note that two scientists can look at the same data and arrive at two different but correct explanations may be attributed with a dynamic view of science. Yet, when viewed from the Kuhn et al. framework, there are two possible epistemological views that might underlie this position: a multiplist or an evaluativist perspective. This distinction has important implications for learning in the disciplines, and most notably in inquiry-based science, because of its impact on an individual's inclination to critique and refine explanations.

Being inclined to judge the value of a claim or explanation rests on a consideration of knowledge as a product of construction. If knowledge is considered immutable, then it makes more sense to simply assimilate the new information, than to question the strength of its warrants or explanatory power (Songer & Linn, 1991). This depicts an absolutist perspective, and this depiction is a point of convergence between the prevalent approaches in the learning sciences and in developmental psychology. Yet, even if knowledge is considered constructed, there may not be much motivation to critique it if radical relativity, multiplicity, is espoused. That is, if people believe that all claims are equally valid then critique may seem like a pointless exercise. This is in contrast to scientific practice, where explanations are continually critiqued and refined in order to arrive at the most parsimonious explanation that explains the widest range of data or phenomena. This type of practice is motivated by seeing knowledge as constructed but subject to adjudication. Or, in other words, on adopting an evaluativist perspective. In sum, an evaluativist, but not a multiplist perspective would be commensurate with science learning goals, but this distinction would not be perceptible in the absence of the Kuhn et al. lens.

## The Present Study

We conducted the present study as a follow up to earlier surprising findings. In an earlier study we compared epistemological perspectives between a “regular” school and a science-centered school that has an extended project and inquiry oriented science curriculum. We had expected students in the science-centered school to tend more toward evaluativism, especially in science. However, these students actually tended toward absolutist views in science (citations omitted for blind review). This led us to consider why an emphasis on science might lead to absolutist views. We conjectured that the emphasis on science made broader societal beliefs, such as the belief in the superiority and certainty of science more salient, and that these broader scripts shaped learners' views about science. In response to our surprising findings, we also began to explore learning in other disciplines, and further conjectured that instruction as it is typically carried out in the humanities, and especially in art, would foster evaluativist views to a greater extent than typical science education. There is a measure of irony in this line of conjecture – in the sense that the disciplines that are less privileged in society, such as art in comparison to science cultivate the epistemological views that are more valued (at least from a Western social science perspective).

Therefore, we set out to compare the epistemological views of students in a scientific discipline and in a humanities discipline. We focused on biology and history as a function of access to participants, and of our prior research experience. We expected history majors to tend more toward evaluativism than science majors. We expected students to have differentiated views concerning their own and other disciplines. As we noted in the introduction, prior research has documented discipline-specific and differentiated epistemological views (Hofer, 2006). Many of these studies employed questionnaires that ask participants about how knowledge claims are handled in different disciplines. We were interested in examining how learners' epistemological views come into play and guide their decisions in contextualized reasoning tasks. Questions about a discipline might trigger stereotypes and societal scripts, but performance on a task, we thought, was less susceptible to these threats. So, we decided to compare students' views by comparing their reactions through structured interviews to competing accounts in their major discipline, in the other group's major discipline, and in an everyday context.

## Methods

### Participants

We interviewed ten university biology majors and ten university history majors in a university in Israel. The students were in their second or third years of study. Undergraduate studies in Israel in the humanities, social sciences and natural sciences are three years long, so third year students in particular were at the final stages of their undergraduate disciplinary education. Importantly, undergraduate studies in Israeli universities are very discipline-focused, that is, almost all of the coursework in completed within the department of their major. This often includes various "service courses," for example, students majoring in education, psychology and sociology may each take a statistics course, but they will not take the course in the statistics department, rather they will take the course in their respective departments. This lends more support to attributing any differences found between biology and history majors to their disciplinary socialization. It also suggests that any disciplinary socialization that we may find may be stronger than in more interdisciplinary contexts (such as the emphasis on a rounded education in liberal arts that is found in many colleges in the United States).

### Task

Our task draws on earlier work by Kuhn and Weinstock (2002), who presented participants with competing accounts of historical cases. Our task included three scenarios from three different disciplines: biology, judicial, history. The biology case was adapted from learning environment design work by Linn and colleagues (Linn, Clark, & Slotta, 2003). The judicial case was loosely adapted from a case used in earlier studies (Kuhn, Weinstock, & Flaton, 1994) to assess argument skill (the original cases were abridged and selective versions of actual cases). The history scenario was adapted from the Kuhn & Weinstock study. . The three disciplines were chosen so that each participant will engage with a scenario from their discipline of study, a scenario from the other group's discipline of study, and a scenario from an everyday domain. Each of the three scenarios depicts competing accounts of the same events. Figure 1 shows an English translation of the biology scenario. Each scenario was followed by a post-task epistemology interview that targeted the participants' views of the dimensions of personal epistemology, mainly the dimensions of certainty, source and justification of knowledge. Figure 2 shows the post-task epistemology interview guide for the biology task. There were only slight variations in the post-task epistemology interviews to accommodate the different disciplines, for example, whether historians or scientists were named in the questions. We also questioned participants about their learning experiences, and in particular, whether they had encountered discrepant accounts as part of their major area studies. Figure 3 shows this post-task learning experiences interview guide.

Across North America frogs are being found that have major physical deformities. Some frogs have deformed eyes. Others have misshapen or multiple legs -- or they are missing their legs altogether! The following are two brief accounts of why the frogs are being deformed.

A brief account of why the frogs are being deformed
By G. Agmon
Scientist investigating the frogs

North American frogs have been found with deformed legs. The deformed frogs have cysts in the area from which their legs develop. These cysts are caused by parasites. The parasites enter the tadpole early in its development and burrow into the area from which their legs develop. As the tadpole develops into an adult frog, the cysts caused by the parasites interfere with normal development, and the frog develops with leg deformities such as multiple legs or no legs.

In order to test whether the parasites cause the deformities in the frogs, small plastic beads were surgically implanted into tadpoles in the location where cysts were observed in adult frogs -- in the area of the body from which legs develop. Many of the frogs that developed from these tadpoles had multiple legs and feet coming out of the area in which the plastic bead was implanted. Parasites in the water are causing these deformities in the frogs.

A brief account of why the frogs are being deformed
by M. Moyal
Scientist investigating the frogs

In recent years, North American frogs have been found with deformed legs and eyes. Chemicals in the water are causing deformities in the North American frogs. The chemicals in the water come in contact with the tadpoles while they are developing, and this contact causes a reaction that interferes with normal development. As a result, the adult frogs that develop from these tadpoles have deformed body parts.

In order to test whether the suspected chemical was in the water, a sample of water from the area where the frogs live was taken. The water was treated with a substance that is known to cause a reaction with the suspected chemical, and the expected reaction did indeed occur. In order to test whether this chemical can cause the observed deformities, the chemical was applied to normal tadpoles, and the frogs that developed had deformities similar to those observed in the North American frogs.

Figure 1: Biology scenario adapted from Linn et al.

Explain in your own words what is going on with the frogs?
- Can you elaborate?

Are the two scientists' explanations significantly different?
- What are the differences between them?

Is it possible that both explanations are correct?
- If not, why?
- If so, how is it possible?
- Is one of the scientists' explanations more correct?

Can someone be certain about what is causing the deformities in these frogs?
- If so, how?
- If not, why not?

Would a third scientists' explanation concerning the frog's deformity be different from the two explanations that you just read?
- If so, why?

Figure 2: Post-task epistemology interview for the biology scenario

When you think about your studies here in college thus far, what seems important to you?
What stands out?

Have you changed in the way you approach knowledge, thinking or problem solving, not just in your studies? What do you think brought about these changes?
[Probe for types of educational experiences, and for educational contexts – college, high school, etc….]

In your studies, have you ever encountered a situation similar to the task you just did where there are two similar, but somewhat conflicting, accounts or explanations for the same event?
Can you describe that situation?
What did you think about it at the time? [didn't think much of it, seemed natural, found it confusing]
Do you think differently about it now than you did then? How so? Why?

Figure 3: Post-task learning experiences interview

### Procedure

A number of research assistants interviewed the participants. Each scenario followed the same sequence: The research assistant gave the participant a sheet of paper with the scenario and gave the participant a few minutes to read through the scenario. When the participant felt that they had read the scenario carefully, the research assistant interviewed them following the post-task epistemology protocol for that task, in a structured interview style (see Figure 1 for an example of the biology interview). The order of the scenarios varied across participants. After all three scenario interviews, the research assistant continued with the interview following the post-task learning experiences interview (see Figure 2). The sessions were audio recorded.

### Analysis

The interviews were transcribed and coded. Transcripts, devoid of identifying information, such as the subject's major, were segmented into statements. Each statement was coded as exhibiting an absolutist, multiplist or evlautivist perspective, based on the characterizations summarized in Table 1 For example, attributing differences between the discrepant accounts to error or lack of skill would be coded as absolutist, attributing differences to equally valid opinions would be coded as multiplist, and emphasizing interpretation over opinion would be coded as evaluativist. These response-level codes were used to assign an overall epistemological perspective per scenario to each interviewed participant. About 20% of the interviews were coded by a second coder with a high degree of agreement between the coders.

The learning experiences interview was not coded. Rather, after each participant was assigned an overall epistemological view, we grouped the learning interviews by major and epistemological view. We examined the interviews to see whether there were similarities in the learning experiences reports and whether these similarities corresponded to epistemological assignment, and or major.

## Findings and Discussion

The results of the epistemological view assignments are presented in Table 2, a capital letter “X” under the corresponding column denotes the epistemological assignment: A—absolutist, M—multiplist, E—Evaluativist. For some participants it was difficult to make a firm decision about their epistemological assignment, therefore, they were coded as being mid-way between two views. This is indicated in the table by two lower case letters “x” positioned midway between two columns.

We found that both biology and history majors tended to be absolutists in all disciplines, most pronouncedly in the judicial discipline, which was outside both group’s area of study. This is surprising, because prior research suggests that the majority of people maintain multiplist positions (Hofer & Pintrich, 1997; Kuhn et al., 2000). History majors were markedly absolutists in the biology scenario, tended markedly toward evaluativism in the history scenario, and tended markedly toward absolutism in the judicial scenario. Biology majors showed a clear pattern only in the judicial scenario where they were markedly absolutists. They also tended toward absolutism in both the biology and the history scenarios. We speak of these patterns cautiously, because our sample is small so assertions concerning patterns are prohibitive, especially in the absence of statistical tests.

The main finding is that history majors but not biology majors exhibit within their discipline of study an epistemological perspective that is distinguished from their views in the other disciplines. Moreover, the history majors’ view in the history scenario, their discipline of study, tended toward an evaluativist perspective, which is considered to be a more productive view in academic contexts. The difference between history and biology majors in the overall patterns of epistemological assignment, and the difference within the history majors’ assignments in their discipline of study versus other disciplines, suggest that history and biology students have different learning experiences and that these experiences give rise to different epistemological views when contending with discipline-specific content. It also suggests that epistemological views, and evaluativist views specifically, can be cultivated through instruction.

We do not have baseline data for these participants documenting their perspectives prior to their college education, so there remains some question about whether these patterns can be attributed to their college experiences. In addition, one could argue that there is a form of self-selection at play here that conflates the results. That is, it may be that people who maintain certain views tend to pursue certain majors. However, if this were the case we would not expect to see the type of within-major outside-of-major trends that we seem to be seeing for the history majors. Of course, these questions need to be probed more rigorously with a broader study that includes this type of baseline data. Nonetheless, we believe that the findings from the learning experiences interview lends support to our findings and claims, despite these methodological limitations.

Table 2: Epistemological view assignment per scenario per participant

| | Biology Scenario | | | History Scenario | | | Everyday Scenario | | |
|---|---|---|---|---|---|---|---|---|---|
| Major | A | M | E | A | M | E | A | M | E |
| History | | | X | | x | $x_a$ | X | | |
| History | X | | | | X | | | X | |
| History | X | | | | x | x | X | | |
| History | X | | | | x | x | x | x | |
| History | X | | | | x | x | * | * | * |
| History | X | | | | x | x | x | x | |
| History | * | * | * | | x | x | X | | |
| History | | | X | | | X | | x | x |
| History | * | * | * | * | * | * | x | x | |
| History | X | | | X | | | x | x | |
| Biology | | | X | X | | | X | | |
| Biology | X | | | | | X | X | | |
| Biology | X | | | X | | | X | | |
| Biology | | X | | x | x | | X | | |
| Biology | | | X | X | x | x | | | X |
| Biology | * | * | * | X | | | X | | |
| Biology | X | | | | | | X | | |
| Biology | X | | | | | | X | | |
| Biology | * | * | * | * | * | * | * | * | * |
| Biology | X | | | | | X | | X | |

The participants' responses to questions concerning their educational experiences at the university also show distinctions between history and biology majors. All of the history majors noted that they had encountered situations similar to those presented in the scenarios, where the same phenomenon or event is explained through competing accounts. Only four biology majors reported encountering competing accounts in their classes.

The History majors responded immediately, assuredly and fluently that they had encountered competing accounts in their studies, and easily provided specific examples. They seemed to associate this idea of competing or multiple accounts with "what History is."

For example:

> "Ah, actually yes. There are a lot of cases like that that show you some event and then ((two undecipherable words)) another direction and then you need uh to decide uh which interpreter you go for more. I usually just combine the two interpretations. Present uh there are some that say this and there are some that say that and I try uh not not to say anything conclusive. Which explanation is more acceptable. Because some think this and some think that. History is not an exact science. That is you can't uh say what is right and what is wrong. Anyone can come with his interpretation." [History major, translated from Hebrew]

> "(Laughing) of course I did. That's what you do in history. (Interviewer: Can you describe the situation). Yes, if I need to hand in a paper on two different sources that I need to compare and check who wrote from what perspective did he write, from what era did he write, and who paid him to write…so like I need to think about these things". [History major, translated from Hebrew]

The Biology majors hesitated, and were hard pressed to provide examples. Some biology majors referred to a different concept, such as being presented with different accounts on an exam and having to identify the correct one (i.e., multiple choice questions), or noted that there are proteins that have two different functions.

Despite their familiarity with the notion of competing accounts, the history majors did not express ideas concerning principled ways of critiquing competing accounts, and for favoring one account over the other. Most of the history students noted that individuals simply have to choose for themselves what seems to suit them most. For example (sic, translated quotes): "…if one book strengthens your view and another that weakens your view, then you will obviously go with the one that strengthens your view…" or "I hold on to what works for me, what I believe in and that's that. I know that both are uncertain." In this respect, the history majors reflected more of a multiplist than evaluativist perspective.

Our findings are commensurate with our expectations that instruction can play an important role in cultivating epistemological views, and that typical instruction in the arts and humanities is more focused on multiplist and evaluativist views than typical science instruction. This is ironic, because as we have noted in the background section, it is the evaluativist perspective that is most commensurate with science learning goals. The overall tendency toward absolutist perspectives is rather disappointing. It is also somewhat disappointing that the history majors seem to lack a disposition to critique, which we posit would be part and parcel of a full and complex evaluativist perspective.

## Conclusion

This study was exploratory. It served mostly to point to profitable research directions, than to provide firm ground for conclusions. We studied a small number of participants, and employed only a single measure. In addition, our knowledge of the instruction underlying the views we explored is limited to narrow self-reports. Our goal is to follow this initial study with a broader investigation. We intend to combine written instruments administered at a large scale with smaller scale interviews, and to complement student self-reports with instructor self-reports and classroom observations. We expect that such a study would enable us to make stronger claims about disciplinary differences in the cultivation and adoption of epistemological views, as well as point to the particular instructional strategies that seem to foster these views. Our hope is that these endeavors will help us understand how instruction and learning environments can advance the refined development of evaluativist perspectives.

## References


Hofer, B. K. (2004). Exploring the dimensions of personal epistemology in differing classroom contexts: Student interpretations during the first year of college. *Contemporary Educational Psychology, 29*, 129-163.

Hofer, B. K. (2006). Domain specificity of personal epistemology: Resolved questions, persistent issues, new models. *International Journal of Educational Research, 45*(1-2), 85-95.

Hofer, B. K., & Pintrich, P. R. (1997). The Development of Epistemological Theories: Beliefs about Knowledge and Knowing and their Relation to Learning. *Review of Educational Research, 67*(1), 88-140.

Hogan, K. (2000). Exploring a Process View of Students' Knowledge about the Nature of Science. *Science Education, 84*, 51-70.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15*, 309-328.

Kuhn, D., & Weinstock, M. P. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 121–144). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education, 87*(4), 517-538.

Schauble, L., Glaser, R., Duschl, R. A., Shulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *Journal of the Learning Sciences, 4*(2), 131-166.

Schommer, M., Calvert, C., Gariglietti, G., & Bajaj, A. (1997). The development of epistemological beliefs among secondary students: a longitudinal study. *Journal of Educational Psychology, 89*, 37-40.

Smith, C., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth-Grade Students' Epistemologies of Science: The Impact of School Science Experiences on Epistemological Development. *Cognition & Instruction, 18*(3), 349-422.

Songer, N. B., & Linn, M. C. (1991). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching, 28*, 761-784.

Tabak, I., & Weinstock, M. P. (2005). Knowledge is knowledge is knowledge? The relationship between personal and scientific epistemologies. *Canadian Journal for Science, Mathematics and Technology Education, 5*(3), 307-328.

# Fostering Mathematical Inquiry: Focus on Teacher's Interventions

Mara V. Martinez and Wenjuan Li
University of Illinois at Chicago

Abstract: Previous research has emphasized the need to better understand and articulate the demands in the work of teaching mathematics entailed by an inquiry-based approach. It is in this context that, first, we describe an inquiry task intended to provide high school students the opportunity to construct algebraic proofs. Second, as students work on the problem, we map students' inquiry process. More, we identify elements common to students' inquiry process and current views of how mathematics knowledge is constructed. Last, we illustrate the teacher's interventions intended to sustain students' inquiry. Among these, we identified: (1) helping students re-focus their inquiry, (2) helping students select mathematical tools, (3) accepting students' provisory ideas, (4) recognizing the potential in students' ideas and promoting the student to showcase the idea, and (5) reviewing a property using an additional example to preserve the original challenge for students.

## Introduction

In the United States, the mathematics instruction that most students experience in today's classrooms embodies myths that misrepresent the nature of mathematics as well as what it means to learn and do mathematics. Mathematics textbooks, pedagogical practices, and current assessment policies work in concert to perpetuate the idea that mathematics is the discipline of certainty. This is despite current reform efforts as voiced in the NCTM Standards (National Council of Teachers of Mathematics, 2000) and as articulated in standards-based NSF-funded curricula. These documents advocate for a shift in emphasis from routine skills and factual knowledge to, posing and solving a variety of math related problems, reasoning and communicating mathematically, and appreciating the value of mathematics. Mathematical inquiry, also known as inquiry-based learning, discovery learning, or inquiry-based teaching, is an approach to teaching and learning that intends to engage students in authentic mathematical activity for the purpose of learning about what is entailed in doing mathematics. However, creating a classroom atmosphere and designing instruction to promote mathematical inquiry presents many challenges for the teacher (Goldman, Radinsky, Tozer, & Wink, in press; Lampert, 1995). For instance, individual students are interested in different questions or taking different approaches to the problems, but the teacher must teach the class as a whole. Or, students are inclined to use informal language or everyday life language in the inquiry process, but using formal language or precise disciplinary language is a major goal of learning in discipline. To facilitate inquiry learning, teachers need to know more about students' inquiry process and their thinking in different stages as well as effective scaffolding strategies that connect students' thinking with mathematics.

Therefore, the first goal of this paper is to describe an inquiry task that intends to provide high school students' an opportunity to learn about algebraic proof (Martinez, 2008). Second, to map students' inquiry process. And, last, to illustrate the teacher's intervention intended to sustain students' inquiry. Before addressing the goals, and to contextualize this study, it is necessary to revisit in more detail the nature of inquiry, mathematical inquiry, challenges associated with the implementation of and inquiry based approach, and research questions that still remain open in the field.

## Framework

### Inquiry

*Inquiry* based learning is a pedagogical approach largely embraced across disciplines and goes back in time as far as Socrates (Goldman, et al., in press) "Inquiry learning foregrounds the questions rather than the answer and places the focus on learning in the learner, not in the material that can be transmitted by the teacher" (Goldman, et al., in press). This is in opposition to the more predominant behaviorist view of learning that reduces learning to the acquisition of ready-made facts through listening, memorizing, and practicing. In an inquiry based learning approach, students in addition to learn the discipline, they would also learn habits of mind (e.g., form and pursue questions and the tendency to think critically, among others.)

In agreement with Lampert (1995), *mathematical inquiry* is an approach to curriculum and instruction that gives the teacher the responsibility for introducing content in a way that is illuminated and modified in response to students' questions and ways of thinking about that content. Specifically, the teacher defines the focus of inquiry by posing problems to the class while students take an active part in acquiring knowledge by generating not only answers, but formulations of problems, definitions of the terms of discourse, and analyses of alternative solutions. Such approach assumes a constructivist epistemology (e.g., von Glaserfeld, 1991) while at the same time integrating contributions from supporters of viewing mathematics as a humanistic discipline (e.g., Ernest, 1991). Moreover, it builds upon Dewey's and Peirce's view of knowledge as "a process of inquiry motivated by doubt" (Siegel & Borasi, 1994). Following Borasi (1994), mathematical inquiry calls for highlighting ambiguity and uncertainty in the mathematical content studied so as to generate genuine conflict or doubt, and, consequently, the need to pursue inquiry. Indeed, doubt and anomalies, namely something that contradicts our expectations, are considered motors for inquiry.

An inquiry stance has implications in the way that not only learning and teaching are conceptualized, but also mathematical knowledge (Siegel & Borasi, 1994). In fact, mathematical knowledge is fallible and created through a non-linear process in which the generation of hypothesis plays a key role; also, the production of mathematical knowledge is a social process and truth is constructed through rhetorical practices. For instance, as evidenced in Lakatos (1976), *once* the original theorem of Euler's on the characteristic of polyhedra was *proved*, it was challenged by several counter-examples yielding revisions on both premises of the theorem and definitions used in its proof. Clearly, this position is contrary to the view of mathematics as certain, objective, and unproblematic. Absolutist, Logicism, Formalist, Intuitionist and Platonist argue that mathematics is a body of absolute and certain knowledge; the truths of mathematics are universal; and mathematics is discovered but not invented (Ernest, 1998). In an inquiry epistemology, mathematics is regarded as a humanistic and cultural product; mathematics knowledge is fallible; and the results of mathematics are a changeable social product (Ernest, 1998).

## Inquiry in the Mathematics Classroom

School mathematics is often presented as a body of absolute and certain knowledge. Learning mathematics in school is about knowing facts and procedures, and getting the right answers as quickly as possible (Lampert, 1990). Richards (1991) argued that the problems in school mathematics are habitual and unreflective and the discourse of school mathematics is just "number talk". On the contrary, mathematical inquiry is about encouraging students to ask mathematical questions; to solve mathematical problems that are new to them; to propose conjectures; and to be active participants in the construction of mathematical arguments (Richards, 1991). Central to the inquiry classroom is to offer students the opportunity to experience the uncertainty in the mathematical content, the nonlinear development in constructing mathematical knowledge, and to generate questions and to justify their answers (Siegel & Borasi, 1994).

Research centered on student thinking (Lampert, 1995) suggests that in this type of environment, the teacher's job should be a relatively simple matter of refining and connecting informal understanding with what we want students to learn in school. Research conducted on inquiry-based classrooms focusing on the work of teaching indicates otherwise.

Goldman et al. (in press) identify three challenges related to the classroom instructional context and the teacher's guidance. First, the teacher's level of preparedness to guide inquiry-learning projects; indeed, effective guidance of inquiry activities requires that teachers understand both the process and the content of the inquiry. Second, teachers need to shift norms of classroom discourse away from the traditional pattern (teacher initiation, student response, and teacher evaluation) to shared construction of reasoned arguments. Third, teachers need to move students' everyday language to a more formal language as it is used in the discipline.

Lampert (1995) identifies several tensions that a teacher navigates in an inquiry based mathematics classroom. For instance, individual students are interested in different questions, but the teacher must teach the class as a whole. Also, students' inquiry may push them deeply into one topic, but the teacher is responsible for their knowledge of a broad range of topics. Additionally, students' conjectures push them in different directions through the subject matter, but the teacher is responsible of keeping track of who knows what and what they still need to learn. Further, students may develop idiosyncratic systems for structuring their understanding, but they are also supposed to learn to communicate with a wider community that shares well-established conventions. In an inquiry-based classroom, teachers construct discourse in response to student activity, thus their responsibility is much grater. Therefore, teachers who want to teach in an inquiry based approach need to be able to navigate these tensions.

From a research perspective, little attention has been paid to the challenges of the teaching practice that must be addressed in this type of classroom environment, and how teachers navigate them. Specifically, less attention has been given to "how" to guide the inquiry process once students were engaged in it. It is in this context that this paper focuses on teacher's interventions intended to sustain students' inquiry process when conjecturing and producing an algebraic proof in a high school mathematics class.

# Methodology

## Teaching Experiment

The data was collected in the context of a teaching experiment conducted in a high school in the greater Boston in Massachusetts by the first author of this paper. One of the goals of the teaching experiment was to offer students the opportunity to conjecture in a context other than geometry[1], to use algebra as a tool to prove, and to produce algebraic proofs. It was intended to provide students the opportunity to experience proof as a way to understand "why" a mathematical phenomenon happens. In other words, in constructing the proof students would have the opportunity to access the reasons that make the mathematical statement true.

The teaching experiment consisted of a total of fifteen lessons. All lessons were video taped and all students' written work was collected. In addition, all small group conversations were audio taped. This allowed having an understanding of the class both at a macro level (whole group) and at a micro level (small group and individual students). Also, all students were interviewed individually twice; once half-way-through the teaching experiment and a second time at the end of the teaching experiment.

A group of nine students, who were in 9th and 10th grades, participated throughout the teaching experiment. Students worked in the same group of three for the duration of the fifteen lessons. Lessons were one hour long and were hold in addition to their regular mathematics class. During the fifteen lessons students worked on a set of seventeen *Calendar Algebra Problems* designed by the first author of this paper. The qualitative analysis[2] presented later is based on one of the three groups participating in the teaching experiment. In what follows, a description of the subset of data that is the focus of this paper is presented along with a detailed description of the problem that students solved during the first two lessons of the teaching experiment.

## The Calendar Algebra Problem

As mentioned earlier, students worked on approximately 17 problems as part of their participation in the project. In this paper, data reported in this paper stems from students' work on the first Calendar Algebra Problem (Figure 1).

Problem 1
Part 1: Consider a square of two by two formed by the days of a certain month, as shown below. For example, a square of two by two can be

| | |
|---|---|
| 1 | 2 |
| 8 | 9 |

These squares will be called 2x2 calendar squares. Calculate the difference between the products of the numbers in the extremes of the diagonals.
Find the 2x2 calendar square that gives the biggest outcome. You may use any month of any year that you want.
Part 2: Show and explain why your conjecture is true always.

Figure 1. Problem 1 from the Calendar Algebra Problems.

Students were provided with calendars corresponding to years 2005-2008 accompanying Problem 1-Part 1 (Figure 1). As part of their work on Part 1, students had to analyze the nature of the outcome of the described calculation (subtraction of the cross product). It was expected that students would anticipate some kind of variation in the outcome in relation to the set of days where the operator is applied. This would contradict students' findings as a result of their exploration of the problem. The outcome is always -7 independently from where (i.e., within a month, across months, and across years) the square is located. This contradiction was intended to function as a *motor for inquiry* (Siegel and Borasi, 1994). The ultimate educational goal of Part 1 of Problem 1 was to get students to produce conjectures (correct and incorrect) about the behaviour of the outcome (i.e., number obtained as the result of the subtraction of the cross product) as it relates to the location of the square in the calendar.

After each group of students produced their conjectures, Part 2 of Problem 1, students had to gather evidence to show that their conjecture was indeed true. In addition, students had to figure out *why* this phenomenon happens, and whether this is *always* going to be the case.

## Data Analysis

Data was analyzed qualitatively taking a grounded theory approach (Glaser & Strauss, 1967), which is a bottom-up approach. In other words, starting from the data theoretical relationships and categories are constructed. As mentioned above, in this paper, a map of students' inquiry process is generated based on the work on problem 1 of one group of three high school students. In addition, among the teacher's interventions we identified and analyzed the interventions that were intended to sustain students' inquiry process.

## RESULTS

### The non-linearity of students' inquiry process

Analysis of the group's discussions revealed the non-linearity of students' inquiry process. According to current epistemological perspectives (Ernest, 1998; Lakatos, 1976; Siegel & Borasi, 1994), mathematical knowledge is constructed through a complex process that involves several stages and includes the production of conjectures, examination of examples and counterexamples, arguments and counterarguments, reformulation of conjectures, redefinitions of terms, evaluation of theories, among others. This is in contrast with a more traditional epistemological view where the mathematician constructs a conjecture, proves the theorem, and as a consequence, truth is established forever. This is what we call linear process. It is characterized by a lack of loops going back to review or question prior accepted knowledge.

Indeed, students went through the following stages: Interpretation of the problem, conjecturing process, agreement on the conjecture to prove, determining how to prove the conjecture, proving the conjecture using algebra, and finally evaluating the implications of their work onto their mathematical knowledge.

Students started exploring the problem by using specific examples. Students will do this by placing the 2x2-calendar square in different places within a month, and in different months. As a result, they conjectured that the outcome would be always -7. This was in contradiction with their expectations. Initially, they had assumed that the outcome would vary depending on the location of the square. This has been described in the literature as follows "most often this feeling of doubt arises when an anomaly- something that doesn't make sense in light of existing beliefs- is encountered" (Siegel & Borasi, 1994, p.222). In this episode students encountered an anomaly, something that contradicts their expectations. This anomaly functions a motor for inquiry.

In what follows, students justified their conjecture stating: "it is -7 because there are 7 days in a week". Students are aware that this is not a proof; however, it seems that the fact that a week has seven days warrants the construction of their conjecture, and in a way, is reassuring. This is what Lakatos (1976) refer to as "conscious guessing" in the process of production of mathematical knowledge. Following Siegel and Borassi (1994), this "guess" enables to set in motion the process of knowledge production.

Once the proving process was set in motion, they seemed to agree that they needed to produce a general argument, probably using algebra. In doing so, students tried different paths, some of them correct and some incorrect. One of the incorrect attempts involved using a linear function when it was not the correct mathematical tool to use. At this point, the teacher stops by the group, and without questioning directly the correctness of the strategy, helps them to shift direction (Episode 1 in section below). The teacher's intervention helped them to make explicit and gravitate towards the idea of using variables so that they can continue working on it.

The next challenge that students faced was how many, and which independent variables to consider to have a complete algebraic model of the situation. First they used two independent variables to represent the numbers in the 2x2-calendar-square when, indeed one independent variable is needed. Once again, the teacher stopped by and helped students' to continue their inquiry process (Episode 2 in section below).

After that, students wrote the expression $a(a+8)-(a+1)(a+7)$ to represent in a general way the calculation performed on the 2x2-calmdar-square. Students were able to distribute transforming the expression $a(a+8)$ into $a^2+8a$. However, things did not go as well with the negative sign of the second term, namely $-(a+1)(a+7)$. At this point, the teacher intervened by recalling an example to illustrate how the negative sign would impact the expression (Episode 3 in section below). After overcoming this challenge, students constructed and algebraic proof showing why the outcome is always -7 independently of the location of the 2x2-calendar-square.

Students' inquiry describes a non-linear process, or what Lakatos called "a zig-zag" path. In this process, students conjectured, revised conjectures, gathered evidence, formulated and reformulated their ideas, devised strategies and re-evaluated them based on feedback provided either by the teacher, their findings, or peers. Students' inquiry process resembles current views of the production of mathematical knowledge. Contradiction functioned as a motor to engage students in inquiry. Students structured their work base on a "conscious guess" and produced mathematical knowledge upon it. In what follows, teacher's interventions are analyzed in the light of scaffolding or sustaining students' inquiry process.

## Teacher's interventions intended to sustain students' inquiry

Teacher's interventions play a key role in sustaining students' inquiry process. However, little attention has been paid to the role of the teacher in sustaining students' inquiry once they are engaged in it (Chazan & Ball, 1999; Lampert, 1995). Given that teacher's discourse is constructed in response to student activity, the teacher's responsibilities and intellectual demands are substantial. The analysis yielded five teacher's interventions intended to sustain students' inquiry process. In what follows, these interventions are identified and illustrated.

### Episode 1: Helping students to re-focus their inquiry

1 Student 2 (S2): It's … these the numbers increase at a constant rate right
2 S2: That's why it's always -7.
3 Teacher (T): So your hypothesis is that it doesn't matter where you place the square, you are going always to obtain minus 7, that is you hypothesis? Ok. So, how do you gather evidence to prove that? That is your problem now. How did you do? How did you come to that idea [outcome is always -7]?
4 Student 1 (S1): Did examples.

Interpretation: As part of their inquiry, students analyzed the underlying mathematical structure of the calendar, concluding that it was linear. Even though this is correct and potentially useful, the teacher interpreted that the students were focusing mostly on the writing of the linear relation. In addition, it seemed that students were aware of the link between the existence of a linear relationship and the fact that the outcome is always -7, as evidenced in the dialogue (lines 2 and 3). However, there was no evidence that students were analyzing *how* and *why* having a linear relation connects to their conjecture (i.e., the outcome is always -7). This indicated to the teacher that students probably thought that the mere existence of the linear relation was *sufficient* evidence to prove their conjecture. Therefore, the teacher brought back to the conversation the conjecture and the question "How do you gather evidence to prove that?" At the same time, the teacher wanted to link this last question to how they had arrived to their conjecture. This was an attempt to promote students to continue building on the examples they had used and, perhaps, construct specific linear relations among the elements of the 2x2-calendar-square. Note that the teacher here tries to sustain students' inquiry process by *returning* the question "how do we prove this?" and, by selecting a portion of their work (i.e., examples used to construct their conjecture) that had the potential to continue building the mathematical relations needed to prove.

### Episode 2: Helping students select mathematical tools and accepting students' provisory ideas

1 T: Ok let's try something then. Can I write in this one? So, what we need to do is to gather information here… How can you show that for any number, this [difference between numbers that are the same column in the calendar and in consecutive weeks is 7] is going to be true?
2 S2: Oh I get it, so we do 'x'
3 T: Ok you can try.
4 S2: All right so like different variables.
5 S2: 'a' 'b' 'c' 'd', ok
6 S1: So, it doesn't matter, in a square is always …
7 T: Why did you come up with the idea of using letters?
8 S2: Using variables?
9 T: Yes, variables.
10 S2: Because if you have a formula that shows that they'll always be the same then it will work for anything so actually
11 T: Ok so try to work on that and see whether that gives us something solid to prove.

Interpretation: Until now, students had made great progress by finding out that in all the examples they tried the numbers in the first row of the square are exactly a week apart (i.e., seven days) from the corresponding numbers below (e.g., $\begin{matrix}2 & 3\\ 9 & 10\end{matrix} = \begin{matrix}2 & 3\\ 2+7 & 3+7\end{matrix}$). However, they were not sure how to represent the relationship mathematically capturing all cases, namely they were looking for a general expression. The teacher intervened by asking a question that included the phrase "for any number" (line 1) with the underlying hypothesis that this would help them to connect with algebra. As a result, one of the students suggested to use "x", linking the problem to algebra. The teacher wanted to check whether they understood why "using x" would work (line 7). As evidenced in the student's response "if you have a formula that shows that they'll always be the same then it will work for anything" algebra was being used in a meaningful way. The teacher intervention helped students to identify the mathematical tool that they needed by re-stating what they were trying to do and, deliberately, including the phrase "any number". By helping students to identify a potentially useful mathematical tool, students were able to continue their inquiry process; therefore the teacher's intervention sustained students' inquiry.

In addition, note that when students suggested the use of four distinct independent variables "a, b, c, and d", even though this idea could have been identified by the teacher "as incorrect or incomplete", the teacher accepted it as a provisory idea. This intervention (accepting the idea) gave students the opportunity to students to further refine the idea by themselves as a result perhaps of their later work. More, students' identification of 4 variables indicates progress in terms of their inquiry process; at this point, they do know that they are working with four variables. They still have to figure out that these four variables are indeed related among them (i.e., a, b=a+1, c=a+7, and d=a+8). This type of teacher intervention, accepting students' provisory knowledge, helps sustain students' inquiry process to that extent that gives students the opportunity to refine their "incomplete" ideas further.

## Episode 3: Recognizing the potential in students' ideas and promoting students to showcase their ideas

1 S2: I think that *a* and *b*… *b* is *a plus 1.* So if we make that 'a' plus 1 then we can cancel out both *a's* and that can be 1
2 T: Oh, did you hear his idea? Say it again please.
3 S2: All right so if we change 'b' to 'a' plus 1 cause that's what it's equal to then we can cancel 'a' and then just have 1 here
4 S3: I guess that'd work. Yeah.

Interpretation: One of the students in the group proposed the idea that *b* is indeed *a plus 1*. The teacher recognizing the potential of the idea in advancing students' work, had him re-stating it for the other group members. Thus, this intervention, first recognizing the potential of the student's idea, and second by showcasing it, helps students to advance their inquiry.

## Episode 4: Reviewing a property using an additional example to preserve the original challenge for students

1 T: What is the impact of this minus in the signs of the elements here? Do you remember that?
…
2 T: Ok. […] Let's say that we have this example here -(3+x), if I want to make disappear the parentheses and this minus there do you remember what we do? Do we remember the rule to change the signs?
…
3 T: I will remind it to you, I will remind it to you, don't worry, don't worry. …
…
4 T: Yes? Ok so that is your challenge now. Here [pointing to the minus sign in -($x^2$+8x+7)], how does this negative here impact the signs here? And, in that way you can get rid of the minus here and the parentheses and continue working in order to cancel,

which was your idea.
5 S3: Ok
6 T: Try that challenge now please.

Interpretation: Students did not remember how to distribute the minus sign in the expression "-($x^2$+8x+7)". In order to help them, the teacher, created another example (i.e., -(x+3)) and remind them how the distributive property works. In doing so, she reminded them how to distribute the minus sign purposely using an expression different from "-($x^2$+8x+7)". This intervention did not take away students' opportunity to apply the property once they remembered how it works. Once the teachers intervened, students knew how the property worked and still had the opportunity to do it by themselves.

In summary, the following five teacher interventions intended to sustain students' inquiry were identified: (1) helping students re-focus their inquiry, (2) helping students select mathematical tools, (3) accepting students' provisory ideas, (4) recognizing the potential in students' ideas and promoting the student to showcase the idea, and (5) reviewing a property using an additional example to preserve the original challenge for students. Employing these strategies, the teacher did not take away from students the intellectual challenge of the inquiry and provided support in response to and built upon students' activity. In doing so, the teacher's interventions sustained students' inquiry.

## Conclusion

First, we provided a description of a task designed to foster students' mathematical inquiry. Specifically, it was intended to provide high school students' an occasion to learn about algebraic proof. The Calendar Algebra problem was designed to provide students the opportunity to experience a contradiction between, their expectations, and their findings as result of their exploration of the problem. Students' experienced this contradiction and functioned as a *motor for inquiry* (Siegel and Borasi, 1994). Second, we mapped students' inquiry process. Further, we identified elements of students' inquiry process that are described in current epistemological views of the nature of mathematical knowledge (e.g., non-fallible, it is produced through a non-linear process). Last, we identified and illustrated five teacher interventions (e.g., recognize the potential in students' ideas) that helped students sustain their inquiry. This is relevant to the field, given that little attention has been paid to how teachers guide students' inquiry once they are engaged in it (Chazan & Ball, 1999; Lampert, 1995). Still, we need to know more about how teachers construct and implement assignments that engage students in mathematical inquiry. We also need to know more about teacher's interventions that foster or hinder students' inquiry process. Once the field systematizes these results, they have the potential to inform the preparation of pre-service teachers and develop programs to better support in-service teachers to implement inquiry-based learning in mathematics.

## References


Borasi, R. (1994). Capitalizing on errors as `springboards for inquiry': A teaching experiment. *Journal for Research in Mathematics Education, 25*(2), 166.

Chazan, D., & Ball, D. (1999). Beyond being told not to tell. *For the Learning of Mathematics, 19*, 2-10.

Ernest, P. (1998). *Social constructivism as a philosophy of mathematics*. Albany, NY: State University of New York Press

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*: Aldine de Gruyter New York.

Goldman, S., R., Radinsky, J., Tozer, S., & Wink, D. (in press). Learning as Inquiry. In B. M. E. Baker, and P. Penelope (Ed.), *The International Encyclopedia of Education* (3rd ed.). Oxford: Elsevier.

Lakatos, I. (1976). Proofs and Refutations. The Logic of Mathematical Discovery, edited by J. Worrall, E: Zahar, Cambridge: Cambridge University Press.

Lampert, M. (1990). When the Problem Is Not the Question and the Solution Is Not the Answer: Mathematical Knowing and Teaching. *American Educational Research Journal, 27*(1), 29-63.

Lampert, M. (1995). Managing the tensions in connecting students' inquiry with learning mathematics in school. In D. Perkins, J. Schwartz, M. West & M. Wiske (Eds.), *Software goes to school: teaching for understanding with new technologies* (pp. 213-232). New York: Oxford University Press.

Martinez, M. (2008). *Integrating Algebra and Proof in High School: Student Work With Variables And Parameters On The Calendar Sequence.* Doctoral Dissertation. Tufts University, Medford, MA.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Richards, J. (1991). Mathematical discussions. In E. v. Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 13-51). New York: Kluwer Academic Publishers.

Siegel, M., & Borasi, R. (1994). Demystifying mathematics education through inquiry. In E. Paul (Ed.), *Constructing mathematical knowledge: Epistemology and mathematics education* (Vol. 4, pp. 201-214). Washington, D.C: The Falmer Press.

von Glaserfeld, E. (1991). *Radical constructivism in mathematics education*. Dordrecht, The Netherlands: Kluwer.

---

(1) In the United Sates, proofs appear mostly exclusively in the geometry curriculum in high school.

(2) The analysis was conducted in all groups participating in the teaching experiment. Given space restrictions we illustrate our findings with the analysis of one of the three groups. We chose this group given that the results that we wanted to illustrate appeared all as part of the work of this group.

# The Effectiveness of Reading Comprehension Strategies in High School Science Classrooms

Phillip Herman, University of Pittsburgh, pherman@pitt.edu
Kristen Perkins, Northwestern University, kristen-perkins@northwestern.edu
Martha Hansen, Evanston Township High School, hansenm@eths.k12.il.us
Louis M. Gomez, Kimberley Gomez, University of Pittsburgh
lgomez@pitt.edu, kgomez@pitt.edu

**Abstract:** Reading strategies have been shown to increase comprehension for younger learners. As students move to middle and high school, there is little opportunity to learn about, practice, and apply these strategies in classes, particularly in the content domains like science. Nonetheless, there is a growing national consensus that high school content area teachers need to better integrate discipline-specific reading supports throughout the curriculum. We evaluate the effectiveness of an effort to integrate reading comprehension strategies in biology classrooms. Students' independent proficiency with the strategies predicted science achievement, even when controlling for prior reading achievement. These results provide evidence that strategies are effective and practical in intact science classrooms. We discuss the implications of the findings for the design of content-area literacy instruction. Finally, we describe our efforts to refine the design of the strategy supports based on the empirical results.

## Learning to Read and Reading to Learn

Reading is a complicated cognitive process that is likely the most important competency acquired in all of schooling. As students move from the early elementary years through high school and beyond, the focus changes from 'learning to read' to 'reading to learn.' Many middle and high school students fail to make this transition. That is, many students can decode individual words on a page but are unable to successfully comprehend the sentences and paragraphs that make up academic text. Based on an influential theoretical model of reading comprehension, we could say that these students are unable to build an adequate text base (direct representation of the semantic structures of text) and situational model (integrated reordering of text content with prior knowledge) (Kintsch, 1998), which represent readers' constructive understanding of the meaning of text. It is well known that American students struggle to read academic texts. About 70% of American eighth-grade students performed below the *proficient* level on the National Assessment of Educational Progress reading tests in 2003 (National Center for Educational Statistics, 2005). Students who struggle to read to learn remain at a disadvantage throughout school and later in life as more desired careers require workers to be able to independently read new material, integrate that material with existing models of understanding of a domain, and generate ideas from the new understanding. In the work described here, we detail an intervention that grew out of our multi-year effort to support high school science teachers as they work to increase students' disciplinary learning by focusing on targeted reading comprehension strategies that students use as they engage with high school science texts. We describe the reading comprehension strategies teachers and students used in intact high school biology classrooms, present some of the research base on the use of comprehension-building strategies with older students, and present a rationale for the embedding of reading comprehension strategies in content domains like science. We provide evidence of the effectiveness of the strategies based on performance of students on two assessments of science achievement. Finally, we take initial steps to revise the reading strategies based on the empirical findings. Our work is particularly congruent with the goals of the ICLS conference that stress how discipline-specific learning needs to be accounted for in effective learning interventions.

### Reading Comprehension Strategies in Science

Contemporary models of reading stress that in order to derive meaning from text, readers rely on both text-driven and knowledge-driven processes (Goldman & Rakestraw, 2000) in which they make connections between elements in a text, their understanding of the text, and between the text and prior knowledge (Wittrock, 1990). A reader must actively construct understanding by integrating existing and new knowledge in part through the application of flexible strategies that help foster, monitor, regulate, and maintain comprehension (Alexander & Jetton, 2000). A number of studies have demonstrated that students' understanding of and memory for text can be strengthened through explicitly teaching students multiple comprehension strategies (Pressley, 2000). Some common reading strategies include paraphrasing, self-questioning, reflecting, marking structures of text, summarizing the gist of text, teaching others about a text, etc.

When learned well, strategies have been shown to increase comprehension. Even so, there is evidence that high schools rarely provide opportunities for students to learn and practice effective reading strategies

(Langer, 2001). Part of the problem is that as students move into domain-specific classrooms in high school, content teachers are often underprepared to meaningfully support reading comprehension generally and domain-specific reading in particular (Gomez, Herman, & Gomez, 2007). Several teachers have reported to us that it should primarily be the responsibility of middle school English teachers to get students “ready to learn science.” High school science teachers are under pressure to cover science content standards. Still, science teachers are frequently frustrated by students’ inability to learn from texts, even if they are not sure how to address the problem. We have found that science teachers will often assign reading to students, but then didactically teach the “important content” (Gomez, Herman, & Gomez, 2007). This approach fails to hold students accountable for independently accessing and learning from text. This marginalizing of text in science is extremely unfortunate given how central reading to learn science is to any meaningful definition of scientific literacy (Norris & Phillips, 2003). One’s ability to comprehend science content from reading is important to the work of scientists in the field, as well as critical to promoting scientifically literate citizens. Although there have been influential calls to action to better integrate literacy within the content domains (Biancarosa & Snow, 2006), there remains little practical advice for content-area teachers that is theoretically and empirically grounded in the disciplines. An important goal of our work is to demonstrate that when students have the chance to learn, practice, and apply reading comprehension strategies that are deeply coupled to domain learning goals and domain practice, content learning will increase along with students’ ability to independently read to learn.

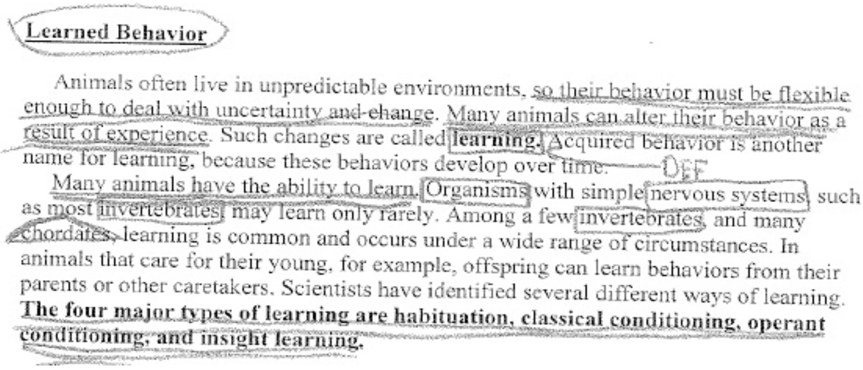
## Reading Strategies: Structure, Reflection and Gist

Science readers should have a corpus of strategies they can use prior to, during, and after reading to learn from text. Students benefit when they are taught to apply comprehension strategies when they read (Anderson, 1992; Collins, 1991). Through repeated transactions with texts and by collaborative analysis and discussion with peers, students can better internalize and ultimately take ownership of reading strategies (Pressley, et al, 1992; Biancarosa & Snow, 2006). When internalized and used frequently, strategy use can lead to large positive effects on text comprehension (Anderson, 1992). The strategies we support (described in detail below) are designed to help students identify the general structure of texts as well as critical discipline-specific elements such as main and supporting ideas, how scientific arguments are constructed in text, what counts as scientific evidence, etc. (Gomez, Herman, & Gomez, 2007). In addition, the strategies should help students know how to reflect on, analyze, and organize text so that elements can be examined and critiqued for understanding and communication (Herman et al, 2008). Finally, students should also know how to summarize a text to integrate new and prior knowledge into one holistic representation of their understanding. This summarization strategy helps students communicate their understanding of the gist of what they have read, and to make the connections between new and prior knowledge explicit (Kintsch, 1998).

The reading strategies implemented in these science classrooms are intended to increase reading comprehension and science learning by developing students’ metacognitive reading skills to increase active and conscious processing of text. Though there are many conceptualizations of metacognition in the literature, we focus on developing students’ conscious control of their cognitive processes (i.e., self-regulation) (Pressley, 2000), including planning, selecting, and using appropriate strategies; monitoring reading comprehension; analyzing the effectiveness of reading strategies; and changing reading behaviors when necessary (Ridly, Shutz, Glanz, & Weinstein, 1992). We conjecture that increases in metacognitive reading skills will allow students to comprehend more challenging text. Over time, as teacher support fades and reading strategies are internalized, students will be able to read challenging texts more independently. Ultimately, these conscious reading strategies will develop into reading skills (Afflerbach, Pearson, & Paris, 2008). Next we describe each of the three classes of reading strategies in detail.

### Annotation

Text annotation is a strategy to make an author’s message more explicit to the reader. Students are taught how to identify and mark important information, and disregard irrelevant information. Students typically annotate (by marking on the text) one or more text elements such as difficult vocabulary words and embedded definitions; main ideas/arguments and related supporting ideas/evidence; and headings, transitional words, and other signposts. Initially, teachers explicitly model this annotation process, but over time reduce their scaffolding for annotation so that students can independently annotate texts. Figure 1 is an example of a student’s annotation from one of the two texts used in the study described later.



**Learned Behavior**

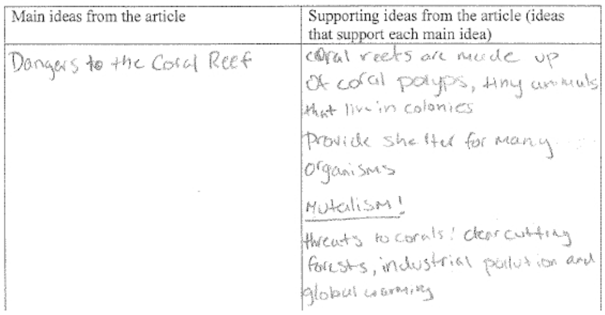
Animals often live in unpredictable environments, so their behavior must be flexible enough to deal with uncertainty and change. Many animals can alter their behavior as a result of experience. Such changes are called **learning.** Acquired behavior is another name for learning, because these behaviors develop over time. DEF

Many animals have the ability to learn. Organisms with simple nervous systems, such as most invertebrates, may learn only rarely. Among a few invertebrates, and many chordates, learning is common and occurs under a wide range of circumstances. In animals that care for their young, for example, offspring can learn behaviors from their parents or other caretakers. Scientists have identified several different ways of learning. **The four major types of learning are habituation, classical conditioning, operant conditioning, and insight learning.**

Figure 1. Example of a Student's Annotation.

## Double-Entry Journals

A double-entry journal (DEJ) (See Figure 2) is a reader-response log that provides a structure for students to monitor and document their understanding of science texts. The DEJ provides students with an organizational tool that suggests corresponding categories of information that students extract from the text, rearrange, paraphrase, and use to reflect on their understanding. The variety of DEJ structures allows teachers to focus students' reading on an important idea or skill that is particularly relevant for a given text (vocabulary, main ideas with supporting ideas, relating information in the text to prior knowledge, etc.), thus coupling the DEJ (and students' attention) to the targeted content learning goal.



| Main ideas from the article | Supporting ideas from the article (ideas that support each main idea) |
|---|---|
| Dangers to the Coral Reef | coral reefs are made up of coral polyps, tiny animals that live in colonies<br>Provide shelter for many organisms<br>Mutalism!<br>threats to corals: clear cutting forests, industrial pollution and global warming |

Figure 2. Example of a Double-Entry Journal.

## Summarization

Summary writing is a critical scientific skill. It requires the reader to effectively digest new information and communicate it in a way that makes sense to her as well as to an external audience. Summarization is a particularly difficult task when students are trying to make sense of texts that exceed their nominal reading level, as is the case for many high school students. In summarizing, students must comprehend the text, identify main ideas, differentiate secondary ideas, integrate new knowledge with prior knowledge, and condense the information in a succinct and logical way.

## Reading Strategies, Science Learning, and Scientific Literacy

Being a competent science learner requires students to learn about science from texts in many forms, including readings, graphs, charts, and other representations of phenomena, processes, and data. Reading to learn science, though certainly not the only competency that accounts for success in school science learning, is nonetheless a constitutive element of students' success in high school classrooms and beyond. In fact, almost any definition of scientific literacy includes the ability to read texts to understand scientific phenomena and thought (Norris & Phillips, 2003). Teachers and students make the strategies science-specific by the tasks they engage in with the strategies. For example, though a double-entry journal could be used in a History class to analyze the causes of a particular war, it is used in science to help students better understand how good scientific arguments require explicit forms of evidence. In this sense, though the *forms* of the strategies are generic, our work has been to deeply couple them *in function* to science learning objectives and scientific literacy.

# The Present Study

Few studies to date have focused on evaluating an ongoing program whose goal is to boost science achievement by focusing on science-specific applications of reading comprehension strategies in intact science classrooms. Most studies of reading strategies have been episodic and somewhat uncoupled from classroom learning goals. Furthermore, few studies have taken place at the high school level (Cromley & Azevedo, 2007), and almost no studies have linked strategy use to domain learning. Strategy interventions have not typically been designed to be discipline specific but rather generic, in the sense of being applicable across disciplines. In this work, we also want to provide an empirical (not just theoretical) base for the design of strategy supports for learners. We want evidence that the annotation, DEJ, and summary activities that we prescribe are effective in increasing science learning. To do so requires revision of the design of strategies, which were initially based on theoretical models of what students should attend to in readings. An important goal of this work is to better understand the connection (both theoretically and empirically) between reading comprehension and domain achievement. Our approach is to support science teachers as they contextualize the reading comprehension strategies in the discipline of biology.

## Research Questions

1. Is proficiency with the reading strategies related to science achievement?
2. Does proficiency with the reading strategies predict achievement above and beyond on-entry reading ability?
3. Can the empirical results in this study inform the refinement of science-specific reading strategy design?

## Participants

All of the work reported here took place during the 2008-2009 school year at a large suburban high school abutting Chicago, IL. Approximately 2,950 students attend the school with 48% being Caucasian, 36% Black, 11% Hispanic. Thirty-four percent of the students are classified as low-income based on eligibility for free or reduced lunch. Approximately 70% of graduates in recent years attended college. In 2005, the school adopted the broad goal of integrating literacy across the content areas. Starting in 2007 administrators and instructional leaders (including a co-author who is chair of the science department) met with our research team to develop a literacy initiative specifically for the science department. Based on what we learned from our prior work focused on embedding literacy in science classrooms (Herman et al, 2008), we worked with a pilot group of biology teachers to introduce ways to couple literacy activities to science learning goals. During the 2008-2009 school year, the intervention expanded to every biology classroom. All biology teachers were provided with ongoing professional development and coaching (by a co-author who had been a high school science teacher at an earlier research site). Fifteen science teachers and approximately 860 students participated in the study.

## Design of the Assessment

Teachers and students in all participating classrooms used the three reading strategies throughout the fall of the 2008-2009 school year (Annotation, Double-Entry Journaling, and Summarization). In January and June of 2009, we administered a one-day assessment of strategy proficiency and science achievement. 752 students were present on both dates and completed both sets of assessments. The design of this assessment was informed by our prior work (Herman et al, 2008), in which we used a point-in-time assessment to gauge how well students were independently able to use the literacy strategies while reading science texts, and the resultant effect on science achievement. It is worth stressing that the assessments described here measure both students' *independent* understanding of, and proficiency with, the reading strategies as well as their *independent* understanding of the science content in the readings. The format of the January and June assessments was identical: Students were randomly assigned one of two readings ("Learned Behavior" or "Coral Reefs") and randomly assigned (within classrooms) to use one of the three strategies while reading the text. Finally, students completed a 10 item multiple choice science assessment that was based on the content in each of the readings. Students used the same class of strategy during both assessment days but the reading and science assessments were crossed so that if a student completed an annotation of the Behavior article in January, they would be assigned to complete an annotation of the Coral article in June. Though we designed and administered two assessments four months apart in order to understand something about growth trajectories in strategy acquisition, that is not the focus of this paper and space limits prevent us from presenting the longitudinal analysis. Instead, the assessment results are grouped together for the analyses presented here.

## Coral Reefs and Learned Behavior

The two readings were chosen from the students' biology textbook. They represent two genres of science texts that students are exposed to: The Learned Behavior text is a more traditional expository reading, whereas Coral Reef is a magazine-type article, representative of the type of chapter inserts that are included in the textbook. In addition to genre differences, the texts were chosen because neither of those topics would be covered in the science classrooms for the duration of the study, thus reducing the likelihood of prior knowledge effects on the science achievement assessment.

## Science Achievement Assessment

For each reading, 10 multiple-choice items were developed. To develop the candidate items, we first analyzed some of the science learning objectives for students that were encapsulated in the content of the texts. We then used those science learning objectives to generate individual items that would assess students understanding of the science content. The construction of this assessment was disciplined by two constraints: 1) As much as possible, we did not want prior knowledge to allow students to answer any of the items so as to better couple success on the items to successful reading of the passage, and 2) We wanted classes of items that varied in difficulty and we wanted that difficulty to be approximately the same across passage assessments. To ensure that the effects of prior knowledge were minimized, we designed a measure that listed the major concepts from each reading and asked students to rate their familiarity and understanding using a Likert-type scale. We gave the prior knowledge measure to 54 students, and piloted the candidate items on 98 students who would not be in the classrooms of the study, but who had covered the same biology content the prior year. These students did not have access to the Behavior or Coral readings; they only completed the multiple-choice items. Based on those results, we altered the item stems, the distracters, and the answer choices. Examples of the items that made up the assessment at three levels of difficulty (Recall, Application, Synthesis) are presented below.

1. (Recall) Which of the following is not associated with operant conditioning?
2. (Application) How does a slow growth rate impact the survival of the coral reef?
3. (Synthesis) Coral polyps would be prevented from obtaining food in the presence of:

## Scoring

To measure strategy proficiency, two of the authors developed a rubric for each reading and for each strategy. Thus, for the Coral article there was a separate rubric for Coral annotation, Coral DEJ, and Coral summary. The rubrics were designed to measure those elements of, say, an annotation, that were deemed by the researchers to be important for correctly answering one or more of the 10 science achievement questions. Figure 3 is a list of examples of the elements that were used in the different rubrics.

**Annotation** – Did the student…?

- Box the vocabulary word *mutualism?*
- Double underline the main idea: "Coral reefs make up a natural ocean habitat that is rich in diversity."

**DEJ** – Did the student include…?

- Main idea #1: "The four major types of learning are habituation, classical conditioning, operant conditioning and insight learning."
- Supporting idea #2: "Many animals can alter their behavior as a result of experience. Such changes are called learning."

**Summary** – Did the student…?

- Make a connection between runoff from clear-cutting forests and the blocking of light for the habitat?
- Use the vocabulary word *mutualism?*

Figure 3: Example Elements From Scoring Rubrics

Two of the coauthors scored all student strategy work. To ensure the reliability of scores, the two investigators scored 20 identical student work samples to determine the inter-rater reliability of rubric-based scores. When scores were disparate (less than 85% agreement), the scorers talked through their understanding of the rubric and student work. The rubrics were then modified to reflect this shared understanding. The scorers then processed another independent sample of strategy examples, and inter-rater reliability increased to acceptable levels. This reliability check was done for each of the six rubrics.

# Results

*Descriptive statistics* for all measures are in Table 2. The reading percentile is the national percentile rank on a standardized test of reading for each student. This percentile is used as a covariate and represents on-entry reading achievement. The students vary widely in reading ability but the mean of 66.91 indicates on average they are above the 50$^{th}$ percentile nationally on reading. The strategy proficiency scores indicate a wide range of proficiency. For each element, a student may receive a 0 or 1 (except for a few items on the annotations that allowed for a 0, 1, or 2). The strategy proficiency scores represent the total score across all the elements in a particular rubric (i.e., the Behavior annotation). It is worth noting that in some cases not one student was able to receive all possible points on the strategy rubrics. The Coral science achievement assessment was slightly more difficult than the Behavior assessment, even though we made every effort to make the two assessments equally difficult.

*Correlations* were calculated so that we could determine, for all students who completed a Behavior annotation in either January or June, whether their strategy score was correlated with science achievement. Every strategy for both readings was correlated with performance on the science assessment. The correlations were significant ($p<.01$ in all cases). Correlations for Behavior ranged from .23 for Annotation to .42 for both DEJ and Summary. For Coral the range was from .35 for both Annotation and DEJ to .33 for Summary. The correlations are more consistent (within each reading) for DEJ and Summary than they are for Annotation. The Behavior annotation correlation was the smallest (.23) though it was still significant.

*Stepwise Regressions* were performed to determine whether strategy proficiency predicts science achievement even after entering prior reading on a standardized exam as a covariate in Step 1. The results of these regressions are highlighted in Table 3. In each model, strategy proficiency predicted unique variance in science achievement above and beyond what reading alone predicts. Because reading is presumed to be a consistent predictor of performance on most academic outcomes, including science, these results may indicate the value of the reading strategies as a means of raising both science achievement and reading comprehension. In some cases, the amount of variance explained by strategy proficiency above and beyond reading was not large (2% of the variance for Behavior Annotation; while in other cases the effect was larger, as in the case of Behavior Summary (7% of the variance). Because we included a "strong" covariate, these numbers are somewhat encouraging in that they suggest that proficiency with the strategies might be helping students learn more science, even if they are struggling with reading.

*t-tests and ANOVA*: We conducted an analysis of each element of each rubric independently of each other to determine if particular items on the rubric "mattered more" in predicting science achievement. In prior work (Herman et al, 2008) we found an inconsistent relationship between annotation score and achievement. In this study, annotation score did predict achievement. In follow-up analyses to our main study, we essentially conducted a simple differential item analysis. For example, on Behavior Annotation item #3, students would receive either a 1 or 0 depending on whether they identified a main idea in the Behavior reading by underlining it. We analyzed the total science comprehension score to determine if there were differences in that score based on whether students received a 0 or 1 on item 3. We grouped all students into two groups for each item and then compared the science score for each group. We used t-tests when the items were scored 0 or 1 and a One-Way ANOVA when scored 0,1, or 2. Though we realized that by using more than 20 comparisons, we were likely to allow for increased likelihood of Type I errors, we were primarily interested in exploring patterns in the predictive utility of elements of the rubrics. So, for Behavior Annotation, there were 30 items on the original rubric and our empirical analysis indicated that only 8 of those items differentiated students reliably on their total science scores. Once we identified those 8 elements, we recalculated the correlation of strategy proficiency with achievement based on those 8 items instead of the original 30. Though we would expect the correlation to increase because our method necessarily would lead to less variance in the predictors, we wanted to gauge how strong those correlations would be for possible future revisions of the annotation format. In this case, the correlation rose from .19 to .36. In all cases, for both readings and all 6 strategies, the correlations rose with the lowest being .32 and the highest being .43. For each strategy we now have an empirically based subset of items that better discriminates the total science score (not each question on the science test but the mean on the 10 items for each student). The number of revised elements for each strategy compared with the total number of elements for each, is as follows: Behavior Annotation 8/30, DEJ 12/16, Summary 7/12, Coral Annotation 17/27, DEJ 8/18, and Summary 8/18. In each case, a substantially fewer number of items does a better job of predicting the science score.

Table 1: Descriptive Statistics of Variables in the Study

| | N | Possible Range | Actual Range | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Behavior Annotation | 210 | 0-37 | 2-37 | 20.66 | 6.07 |
| Behavior DEJ | 212 | 0-16 | 0-14 | 5.06 | 2.74 |
| Behavior Summary | 187 | 0-12 | 0-9 | 3.91 | 2.30 |
| Coral Annotation | 209 | 0-36 | 1-34 | 15.70 | 7.28 |
| Coral DEJ | 209 | 0-18 | 0-14 | 6.74 | 3.03 |
| Coral Summary | 189 | 0-18 | 1-15 | 6.26 | 2.98 |
| Behavior Assessment | 856 | 0-10 | 0-10 | 6.34 | 2.05 |
| Coral Assessment | 844 | 0-10 | 0-10 | 6.06 | 2.08 |
| Reading Percentile | 856 | 1-100 | 1-100 | 66.91 | 28.01 |

Table 2: Regressions for 6 models: Predicting science achievement from strategy proficiency and reading.

| **Behavior** | Models* | F | Adjusted $R^2$ (Reading only) | Adjusted $R^2$ (Reading + Strategy) |
|---|---|---|---|---|
| Model 1* | Reading Achievement + Annotation | 41.07 | .27 | .29 |
| Model 2* | Reading Achievement + Double-Entry Journal | 63.71 | .37 | .40 |
| Model 3* | Reading Achievement + Summary | 42.33 | .25 | .32 |
| **Coral** | | | | |
| Model 4* | Reading Achievement + Annotation | 36 | .23 | .26 |
| Model 5* | Reading Achievement + Double Entry Journal | 46.65 | .31 | .33 |
| Model 6* | Reading Achievement + Summary | 59.37 | .37 | .39 |

*p<.02 for each predictor in each model

## Discussion

Several findings from this study are worth highlighting. Students who successfully used reading comprehension strategies while reading a science text performed better on a measure of science achievement. That is an important finding for the adolescent literacy and science education research community as well as for practitioners who might wonder about whether the support of reading in high school science classrooms is possible, practical, and useful. This study indicates it is useful to regularly incorporate explicit instruction in reading comprehension strategies in science classrooms, not as a way to primarily improve students' reading abilities, but as a means of increasing science achievement. Science teachers (and other content-area teachers) need better ways to support reading in the disciplines. All too often in science class, readings are ignored or marginalized. Teachers are not to blame for this. Most schools of education provide high school science teachers with almost no preparation for supporting reading in science. Teachers need to know and be able to scaffold a variety of effective reading comprehension strategies that can help increase science learning. This kind of work requires extensive, practice-based, and ongoing professional development.

This study also shows that reading plays an important role in science achievement. The regression models indicate that reading achievement is a significant predictor of science achievement. This finding, though not unexpected, is important because it adds to the understanding of how reading achievement matters for various formats of science assessment, for a variety of learning goals and science text genres. A demonstration of how much reading matters can be a powerful way to spur science teachers to acknowledge and ultimately take some ownership of students' reading abilities in science. We have seen in prior work (Authors, 2007) that science teachers do take ownership over reading when they have a repertoire of teaching strategies that they can utilize and when, as in this case, they can see the significant connection between reading and science achievement. The measure in this study is very similar to other classroom measures of science achievement. We acknowledge that the measure used is not a "pure" measure of science achievement; it is also a measure of reading comprehension. However, we suggest that many classroom assessments of science learning are reading-dependent in meaningful ways.

Student proficiency with any of the strategies was related to student performance on the science assessment. This is worth noting, particularly because that relationship holds even when on-entry reading ability is a covariate in the regression models. Strategy proficiency predicts unique variance in science achievement even when controlling for reading. Based on the adjusted $R^2$ s, reading achievement and strategy proficiency account for between 26% and 40% of science achievement. In prior work (Authors, 2008), we found that proficiency with annotation, though correlated with science achievement, did not predict science achievement when accounting for on-entry reading. We suspect that the reason annotation inconsistently predicts achievement might have to do with the variety of ways that students have learned to annotate. It is possible that only certain elements of annotations matter to comprehension because students may not be actively engaged in completing all the elements of an annotation with all readings. To better understand whether this is true, we conducted some further analyses of our data and identified the subset of elements of the annotation that most differentiated student science achievement. For example, as we mentioned earlier, for the behavior annotation, 8 of the 30 elements significantly differentiated students who did better or worse on the science achievement. The prior research indicated that annotation strategies have utility if they help students recognize and leverage the structures of text for understanding. But, the results of the preliminary analysis presented here indicate that perhaps only the content-focused steps in annotation (like marking main ideas or differentiating specific evidence that supports a particular argument) is important in increasing science learning, and that the more conceptual-focused steps (like circling a heading or subheading) are less useful. There is some evidence that the revised elements are more content focused than conceptually focused. The empirical evidence about strategy efficiency is very important because any intervention that stresses reading comprehension will have to be implemented in real classrooms by busy science teachers. It will require ongoing effort to support schools so they routinely use comprehension-focused, discipline-specific reading strategies in science. Teachers need to be sure they help students learn science. We have provided initial evidence that they do.

Perhaps the major challenge to improve reading in the domains in high school is not discovering new strategies that work but rather researching ways to impact school practice through the kind of research described here. Learning science research can be an important resource in the transformation of science instruction. Much is known about how to teach early elementary reading, and many schools have now put into practice what was only a research agenda a decade or two ago. We do not know nearly as much about high school reading, but we have some good ideas to put into place. Learning scientists need to collaborate closely with schools to make more routine what we know to be effective. Explicit attention to comprehension-focused reading strategies is probably one piece of a good solution to the challenge of supporting reading to learn competencies. This study makes some progress in providing evidence about the utility of strategies, but also goes beyond that by examining how the instructional versions of strategies might be altered by empirical evidence. This could be the most important, if preliminary, contribution of this work. The reading strategies movement was based in large part on a theoretical model of how strategies could help readers. In the case of annotation, it was based on a largely theoretical assertion that readers could better leverage the structures of texts in order to better comprehend text. Good readers seem to better leverage the structure of texts. So, we (and others) designed annotation supports based on that insight. But now we are able, in part by the empirical analysis presented here, to begin to revise and better understand how to link reading activity support to specific science learning. This is an important future direction of this work.

## References

Afflerbach, P, Pearson, P. D., & Paris, S.G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*, 364-373.

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research, Vol. 3* (pp. 269-284), Hillsdale, NJ: Erlbaum.

Anderson, V. (1992). A teacher development project in transactional strategy instruction for teachers of severely reading-disabled adolescents. *Teaching & Teacher Education, 8*, 391–403.

Biancarosa, C., & Snow, C.E. (2006). *Reading Next—A Vision for Action and Research in Middle and High School Literacy: A Report to Carnegie Corporation of New York* (2nd Ed.). Washington, DC: Alliance for Excellent Education.

Collins, C. (1991). Reading instruction that increases thinking abilities. *Journal of Reading, 34,* 510-516.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311–325.

Gomez, L., Herman, P., & Gomez , K. (2007). Integrating text in content-area classes: Better supports for teachers and students. *Voices in Urban Education*, *14*, 22-29.

Herman, P., Gomez, L. M., Gomez, K., Williams, A., & Perkins, K. (2008). Metacognitive support for reading in science classrooms. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 1* (pp. 342–349).

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

National Center for Education Statistics. (2005). The Condition of Education 2005 (NCES 2005–094). Washington, DC: U.S. Government Printing Office.

Norris, S. P., and Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*, 224–240.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research, Vol. 3* (pp. 545-563), Hillsdale, NJ: Erlbaum.

Pressley, M., El-Dinary, P.B., Gaskins, I., Schuder, T., Bergman, J., Almasi, J., et al. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal, 92*, 511–555.

Ridley, D., Schutz, P., Glanz, R., Weinstein, C. (1992). Self-regulated learning: The interactive influence of metacognitive awareness and goal-setting. *Journal of Experimental Education, 60*, 293-306.

Wittrock, M. C. (1990). Generative processes of comprehension. *Educational Psychologist, 24*, 345-376.

# Making Knowledge Building Moves: Toward Cultivating Knowledge Building Communities in Classrooms

Katerine Bielaczyc, Learning Sciences Lab, NIE Singapore, *kateb369@gmail.com*
John Ow, Innova Primary School, Singapore, *john_ow_eu_gene@mac.com*

**Abstract:** A major emphasis within the Learning Sciences has been to explore ways to create classroom cultures that mirror disciplinary cultures. The central focus of our own research has been on how teachers and students navigate the epistemological shift from traditional didactic classrooms to classrooms that function as *knowledge building communities* (Scardamalia, 2002; Scardamalia & Bereiter, 2006). We ground our inquiry in *Ideas First*, a design-based research program focused on creating a knowledge building community in science within a Singapore primary school (Bielaczyc & Ow, 2007; Ow & Bielaczyc, 2007, 2008). In Ideas First we view engagement in the disciplinary practices of science through the theoretical lens of *epistemic games* (Collins & Ferguson, 1993; Morrison & Collins, 1995). In order to support reflective discourse among teachers and students around epistemic game play, we have created a set of dialogic tools centered on specific types of knowledge building moves. We illustrate the design features of the tools and their role in fostering participation in knowledge building communities in science.

Education in the Knowledge Age calls for socializing students into the world of work with knowledge (Bereiter, 2002). A major emphasis within the Learning Sciences has been on investigating ways to create classroom cultures that mirror disciplinary cultures, thus socializing students into working with knowledge in ways consistent with disciplinary norms and practices (e.g., Gresalfi & Cobb, 2006; Herrenkohl, et. al., 1999; Songer, 2006). Such a framing is founded on theoretical perspectives emphasizing learning as a process of enculturation, with a focus on *learning to be* rather than simply *learning about* (Sawyer, 2006; Thomas & Brown, 2007).

In our own research, we have been interested in how teachers and students navigate the epistemological shift from traditional didactic classrooms to classrooms that function as *knowledge building communities* (Scardamalia, 2002; Scardamalia & Bereiter, 2006). For the past several years we have been engaged in a design-based research program in a Singapore primary school. *Ideas First* is a two-year science program co-designed with primary school teachers that has been operating in fifteen grade 3 and grade 4 classrooms since 2006 (Bielaczyc & Ow, 2007; Ow & Bielaczyc, 2007; 2008). The program is based on the vision of a knowledge building community where students work to advance the science understanding of the classroom community through engaging in collectively building knowledge in response to problems of understanding.

Knowledge building communities involve not only participating in knowledge building practices, but also engaging in reflective discourse on participation (Bielaczyc & Collins, 1999; 2006). Such meta-level discourse among participants is critical for agency, identity, and creativity within a dynamic community whose goal is to continually advance at the edges of the community's understanding. In Ideas First we have introduced several means of supporting reflective discourse. In the present paper we describe a particular set of tools meant to support inquiry into specific types of knowledge building moves among both teachers and students. We illustrate the design features of the tools and their role in fostering participation in knowledge building communities in science.

## The Need for Dialogic Tools to Support Inquiry into Knowledge Building

Creating disciplinary cultures in classrooms is challenging because enculturation necessitates an immersive approach. If the classroom currently reflects a traditional didactic culture then change must occur along many dimensions of social and technological infrastructure (e.g., Bielaczyc, 2006; NRC, 2007). When the teachers are themselves new to such a culture, classroom change becomes even more challenging. In Ideas First we chose to work with teachers across all the grade 3 and 4 classes within the school in order to foster a community of teachers that could support each other locally, as well as connect to teachers and other educational stakeholders in the international knowledge building community[1]. In order to foster teachers' development as a knowledge building community and agents of change, as designers we wanted to create supports for critical, reflective discourse --- tools that provided both a lens for looking at the work of the knowledge building community and that helped participants develop a language for talking about it.

Like many designs, we have found that classroom videos provide a powerful means of reflection (e.g., Goldman, et. al., 2007). Further, *Knowledge Forum*, the online environment designed by Scardamalia and Bereiter to support knowledge building (Scardamalia, 2004), also provides a means of visualizing not only students' work with disciplinary content, but also processes of knowledge building. Knowledge Forum records participants' work with ideas using the View feature, a public space where knowledge objects (multi-media content within the database that is represented via icons such as *Notes, Rise-Above's,* and *Build-On's*), their interconnections, and other organizational representations of the knowledge objects (e.g., timelines, maps) are visually displayed. The processes of knowledge building are thus captured in a visual form as the participants' work evolves.

However, while viewing rich visual spaces such as videos and Knowledge Forum where the work of the community occurs in all its "blooming, buzzing confusion" is important, we have found that it can be very difficult for teachers (and students) to "see" the critical events and features (refer to Frederiksen, et. al, 1998 for similar findings in teacher video clubs). Thus we wanted to complement the reflections on classroom videos and Knowledge Forum by creating a set of tools that focused on critical aspects of knowledge building. We thought that a powerful focal point for discussion would be "knowledge building moves," a way of conceiving possible actions to advance knowledge within a given problem space. Further, we wanted to create tools that could be used by *both* teachers and students to engage in inquiry about critical aspects of knowledge building. It should be noted that Knowledge Forum has a growing suite of tools (the Analytic Toolkit) that provide representations of knowledge building activities within the database, thus supporting rich, reflective discourse (Zhang, et. al., 2009). However, because computers are not readily accessible to students and teachers in their classrooms, we constrained our designs to mobile, physical artifacts that students and teachers could easily share.

As stated earlier, we felt that it was critical to work together with the community to develop not only a lens for seeing, but also a language for discussing new concepts. In the early years of the project, when we spoke with teachers about socializing students into working with ideas, it seemed that the notion of "working" was taken as some laborious set of tasks or indicative of routines or procedures. Hence, it became important to find ways to convey work with ideas in knowledge building communities as creative and playful. One useful framework for supporting discourse grounded in playfulness and creativity is *epistemic games* (Collins & Ferguson, 1993; Morrison & Collins, 1995). Using this framework, we have created a set of tools for engaging teachers and students in inquiry into knowledge building moves. We describe these tools in more detail below.

## Knowledge Building Communities and Epistemic Game Play

In creating the theoretical lens of *epistemic games* (Collins & Ferguson, 1993; Morrison & Collins, 1995), Collins and his colleagues are concerned with characterizing the ways in which members of a community of practice work to construct knowledge. *Epistemic games* refer to strategic play with disciplinary knowledge and are based on the study of disciplinary communities such as Physical, Biological, and Social Scientists (e.g., the Cost-Benefit-Analysis Game, the Systems-Dynamics Game). The overall goal is to support learners in developing *epistemic fluency*, developing an understanding of the moves, constraints, and strategies for working with various forms of knowledge.

In Ideas First, we view epistemic game play as being of two major kinds. The first concerns science representational games involving target structures such as classification trees and graphs. The second concerns knowledge building games involving goal states such as the creation of knowledge useful for explaining a problem of understanding. The first kind correspond to those described by Collins and Ferguson (1993) where the game involves constructing a fixed representational target or *epistemic form*, such as playing a Multi-factor-Analysis Game with an And/Or Graph. The second are those where there is no specific epistemic form targeted, such as the Theory-and-Evidence Game (Collins, in press; Morrison & Collins, 1995).

Although we do not use the actual term "epistemic games," we do frame knowledge building in Ideas First in terms of game play and knowledge building moves. Scientists advance their understanding of problems in many ways. One way we represent play with ideas is the "Progressive-Investigation Game" (Figure 1). This representation is meant to capture a type of knowledge building progression that occurs both in the offline setting and online in Knowledge Forum. Carrying the game metaphor further, we discuss with teachers the similarity to sports teams where both full-length games and targeted practice sessions are a continual part of a player's development. Similarly, Ideas First involves a constant cycling of action and reflection in learning to play knowledge building games in science. In order to support teacher and student inquiry into knowledge building moves we have created two specific tools that isolate parts of the full Progressive-Investigation Game for practice and reflection, *Think Cards* and *hypothetical game-configurations*.

Think Cards can be thought of as concrete "game pieces" to facilitate game play (Figure 2). To provide consistency between offline and online game play, the cards are based on Knowledge Forum Notes (Scardamalia, 2004). The Think Cards are used to support one sequence of knowledge building moves that might be used to advance understanding of a problem --- generate an explanation for a problem and then gather new information that can be used to improve this explanation. Think Cards physically reify students' explanations (*My Idea is…*), the new information that they bring to their inquiry (*New information…),* questions that drive their inquiry (*INTU* stands for "I need to understand"), and improvements that they make to their explanations (*A Better Idea is…*).

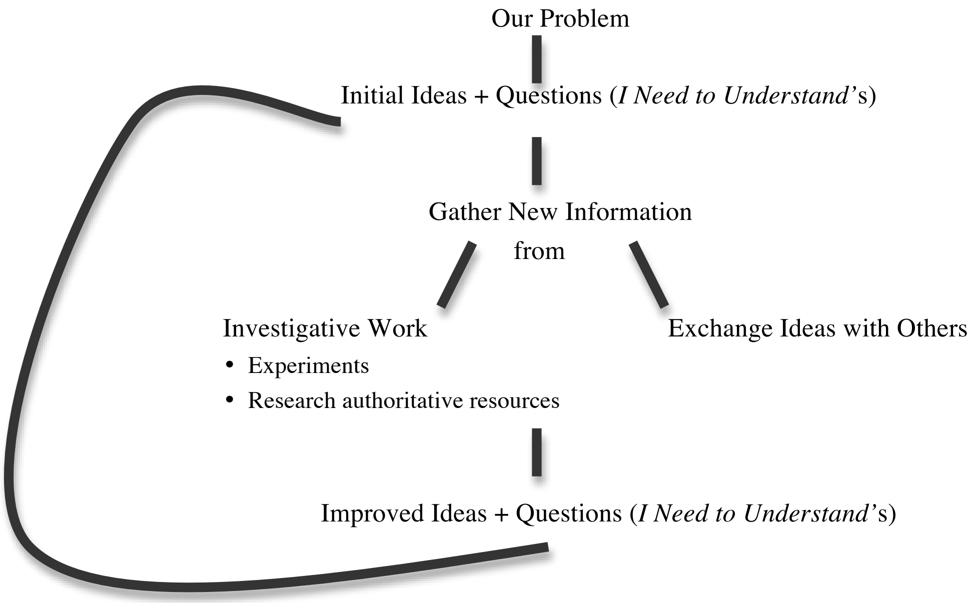



Figure 1. The basic flow of the "Progressive-Investigation Game"

Hypothetical game-configurations are used to reflect on the knowledge building moves made possible by a particular configuration of knowledge objects. The configurations consist of "snapshots" of hypothetical student work in Knowledge Forum, meant to capture game play at a fixed point in time in order to engage the community in asking: given this configuration, what types of knowledge building moves would best contribute to advancing our knowledge? Some configurations focus on single moves, such as presenting a possible initial explanation generated in response to the problem the students are working on. Students then generate a knowledge building move meant to advance this initial idea. There are also more complex game configurations that present not only a possible initial explanation in response to a problem, but also provide a series of possible knowledge building moves. In this case, students both evaluate the quality of the provided moves and generate a possible next move that contributes to the progressive improvement of ideas. In all cases, students each work on the same hypothetical game-configurations so that they can then compare and contrast their proposed knowledge building moves in whole-class discussions about issues such as what makes a "good contribution" and what does it mean to advance the community understanding.

Our design argument is that in full-blown game play in spaces such as Knowledge Forum and classroom discourse, it is difficult for beginning learners to see critical events and features, thus making it difficult to develop the necessary epistemological perspectives on community practices and an understanding of the moves, constraints, and strategies for working with various forms of knowledge (i.e. epistemic fluency). The Think Cards are used in the early months of Ideas First. After students have moved into full-game play in Knowledge Forum, the hypothetical game-configurations are introduced as a means of examining specific elements of epistemic game play, permitting further practice and reflection on participation in community knowledge building. The continued use across grades 3 and 4 makes it possible to highlight particular knowledge building moves such as how to advance an explanation or how to synthesize ideas, which are then related to student work in Knowledge Forum. We provide more detail on our use of the Think Cards as a reflective tool in the Ideas First program in the remainder of the paper.

## Think Cards: Learning to Make Knowledge Building Moves

The Think Cards are used in the first two units of Grade 3 (months 1 - 3). This is students' first formal science experience in primary school. The unit opens with a whole class discussion of "How do Scientists make sense of the world?" and highlights the underpinnings of Ideas First, such as working as a science community to understand questions that we have about the world and how, like Scientists, we keep working to improve our ideas and explanations. The Think Cards are used to support actions consistent with this framing, but are limited to the *initial idea-new information-improved idea* sequence of knowledge building moves. The goal is to challenge the prevailing classroom culture where students' written responses tend to be viewed as static entities that either match a predetermined "model answer" or not. Instead, the first Think Card, *My Idea is…*, is used to encourage students to write down their initial idea, with teachers setting the classroom norm of respect for others' ideas and not worrying if one's idea is correct. These ideas are then shared in the public space (via the Whiteboard, or in small-group discussions) in order to make visible the diversity of ideas that are now available as resources for the whole community. Because of curricular time constraints[2], students work on a problem of understanding that is provided to the whole class (e.g., How do we know if something is a "living thing"?). However students are encouraged to continually record their own questions (the *I Need to Understand…* section at the bottom of the cards) and to use the processes modeled with the class problem to advance their understanding of these questions.

The *New Information…* Card is used to introduce a possible knowledge building move toward progressively improving an initial idea. Rather than using resources to "find the answer," the focus is on using books, videos, the Internet, and other people as resources to improve initial ideas. In many classes, students begin bringing resources from the public library and home to share with the class. Class discussions center on issues such as why certain resources are useful (with some classes discussing the trustworthiness of science content in sources such as children's cartoons and television shows) and the mechanics of note-taking, which is a new skill for these 9 year-olds. Students also discuss the practice of citing resources (including people) in order to return to sources if necessary.

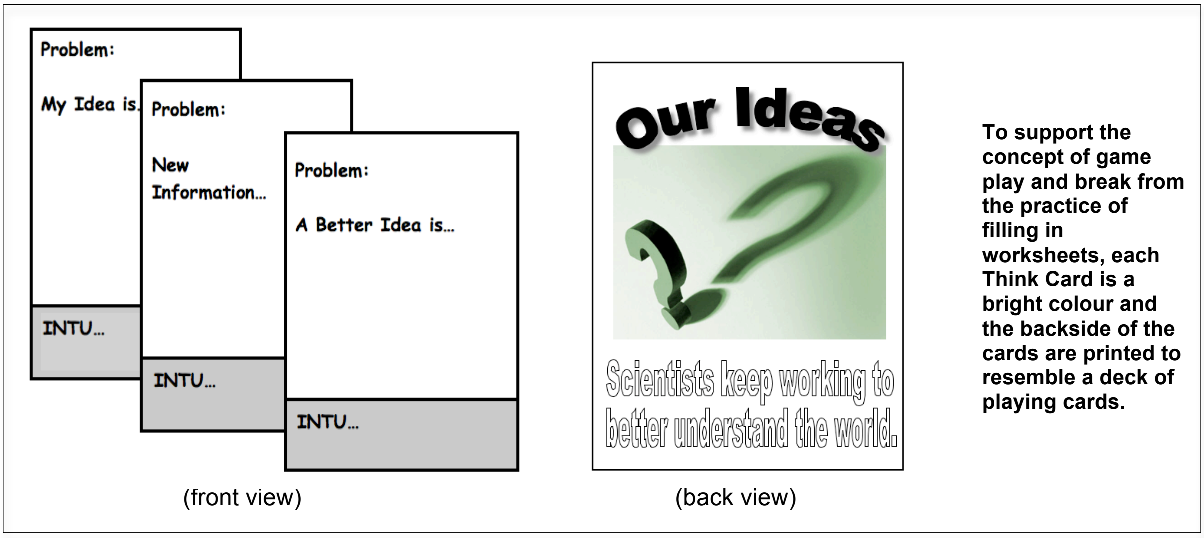



Figure 2. Ideas First Think Cards

The use of the Think Cards occurs over several weeks. The initial idea cards lead into research with various resources to collect new information centered on the class problem, but also including individual questions. Students generate as many New Information Cards as they wish, and share them with the class through whole-class and small-group discussions. This, in turn, leads to access and exchange of information across cards. This work culminates in *A Better Idea is…* Card, where students work to synthesize their learning into an improved explanation of their initial idea. This knowledge building move opens up a space for discourse concerning the meaning of idea improvement. For example, are we interested in looking at improvement in terms of comparison to the initial idea, or solely with regard to the quality of the end product?

The small, mobile nature of the Think Cards make it easy for children to work with their ideas --- jotting down notes in the library or during discussions, spreading their cards out to examine the collection of ideas, literally "exchanging ideas" with others, etc. We believe that it is important that a child can physically accompany the written form of his or her idea into a group discussion, thereby disrupting the conception that a written idea is a static response to a question when the child holding the Think Card is asked by peers to further elaborate the idea or the child defends the idea when it is challenged. The Think Cards also make visible the diversity of ideas that students generate for a particular problem of understanding and that can be collected from various resources, and the multiple pathways possible in moving from initial ideas to construct new knowledge.

Many of the visual affordances (in contrast to the physical) that we discuss here are also available in Knowledge Forum. Such similarity is seen as critical in order to create a consistent developmental trajectory within a semiotic space where knowledge and sources can be questioned and problematised --- from physical knowledge objects in a child's hand that co-located agents can act upon (months 1 - 3 of Ideas First), to "concrete" knowledge objects in the virtual space of Knowledge Forum that can be acted upon by multiple agents (months 4-24 of Ideas First), to broader disciplinary knowledge worked upon by members of the science community.

## Teacher Reflections on the Use of Think Cards

Over the course of the year, teachers in both the Grade 3 and Grade 4 levels meet for bi-weekly reflection sessions. One of the Grade 3 sessions during Unit 1 focused specifically on the use of Think Cards, with the eight teachers all writing written reflections and then sharing their perspectives in an open discussion. Here we briefly illustrate their descriptions of classroom experiences.

One teacher described the Think Cards as a "tool to initiate children into the Ideas First culture," and highlighted that their use "Helps the teacher reinforce the idea that "all ideas are important" by encouraging them to write down their thoughts, whether or not they think they are right." Other descriptions related to culture building included "Students get to experience the processes real Scientists go through in the work," and "*A Better Idea Is…* Card is good as pupils start to rephrase the main content's ideas with vocab and language like scientists. "

The teachers' reflections indicated that they used the Think Cards not only to help students participate in knowledge building practices, but also to engage in reflecting upon the nature of their participation and the ways in which knowledge is built. Based on initial analyses of the teacher responses, two categories emerged concerning the ways the Think Cards helped to support student reflection on knowledge building: *how to engage with ideas and advance understanding* and *how students can serve as resources for each other*. We briefly describe these two areas below.

## How to Engage With Ideas and Advance Understanding

Based on interviews with the teachers of this school over the four years of the project, the typical classroom experience for these students involves answering questions (either verbally or on worksheets) and receiving positive or negative feedback with an explanation. The underlying processes for developing their own explanations are rarely made visible, beyond "studying" or "working harder." Learning that there are processes that can be undertaken to advance one's understanding is thus critical for helping students learn how to engage with knowledge. The teachers describe how the Think Cards introduced students to specific approaches to advancing their ideas and enabled students to see the growth of their ideas:

> *The cards allow pupils to take step-by-step approach in researching their ideas and problems.*
>
> *Helps students to appreciate their own work and observe how they had "grown" in their understanding of a problem.*
>
> *When pupils put their 3 cards together, they can see how their ideas have grown from rather vague and haphazard to something more detailed and specific to the problem or question.*

One teacher told of how using the Think Cards to reflect on their growth of understanding helped her students to "value the process":

> *I saw a shift in their attitudes toward the cards. When we first started out with the Think Cards they were asking me, "Why are we doing this?" And then, um, I had one session with them where they laid out their cards together after they completed one topic. And I got them to sort of like celebrate, or appreciate, the work that they had done individually and at the same time go around and see what their other friends have done. So, um, and to particularly look for how they had grown in their understanding of the problem. So, from there they saw that there was value to what they are doing. So by the time I gave them the second set of cards they quite readily worked without any more of these questioning.*

Through examining their work with Think Cards --- coming to understand that an idea can be a starting place rather than a terminal answer, observing that knowledge can be constructed via multiple pathways, discussing the similarities and difference between what they do and what scientists do --- the intention is for students to better understand not only the processes of knowledge building, but also their agency for participating in such practices (what Scardamalia (2002) refers to as "epistemic agency").

Although the teachers found it helpful that the step-by-step approach of the Think Cards provided an entry point into the processes of knowledge building, the teachers also felt it was important that such an approach did not lead to students following this sequence in a routinised manner, but instead led into more versatility in knowledge building. For example:

> ***Miss H****: I think at this stage when they first just started in this I think they need a little, some steps. So by the second set of cards they are quite comfortable. ...But I think that, um, as they get the hang of this, that the other Thinking Phrases*[3] *can start coming through. They need not follow this, um, steps, steps they have to follow, later on [inaudible] start using other Thinking phrases.*

***Researcher-Facilitator**: So you already foresee a time …*

***Miss H**: Yeah. Because I don't want them to just be thinking this way. First the yellow* [My Idea is card]*, then the purple* [New Information card]*, then the, the, yeah… I think in real life we don't do this, um, all the time. We need them to be very versatile in their thinking as well, not just thinking sequentially but perhaps concurrently with other issues that are popping up.*

It is critical when using supportive processes to guide student introduction to epistemic game play that the processes not be reduced to simple steps to be followed, but instead are viewed as moving students onto a trajectory leading to greater creativity.

## How Students Can Serve as Resources for Each Other

Working together and serving as resources for each other are critical aspects of playing epistemic games in knowledge building communities. With regard to how to serve as resources for each other, the teachers described ways of using the ideas of others in one's own explanations or contributing to each other's advancements, as well as strategies for synthesizing ideas and resolving conflicts. One of the ways they did this was by connecting student use of the cards to the various design principles of Ideas First[4]. For example:

*I strongly feel that the Think Cards help in Principles 3—"Working together to improve ideas."*
- *Students share ideas aloud, teacher writes on board*
- *Other students who do not have the idea will write in their Think Card and see their ideas <u>GROW.</u>*

*It* [Think Cards] *is handy and enables pupils to pass them around to share their ideas.*

*If ideas written on board has misconception, getting second or third student to rephrase the idea promotes Principle No.5—"Moving ahead…and help…"* [The community can only move ahead if we help all of the members of our community to move ahead] *where pupils themselves clarify their peers' idea. If this happen where 2 pupils help one another in clarifying his/her peer's ideas, tr* [teacher] *will write both names in brackets, to acknowledge their contribution.*

Although the Think Cards focus on "My Idea" and "My Better Idea," the teachers indicate that they keep the focus on *our* ideas through keeping ideas visible as public resources for all and through encouraging peers to exchange their ideas and help each other. Such a focus is related to Scardamalia's (2002) notion of "collective cognitive responsibility," a shared responsibility for advancing the community's understanding.

## Supporting Teacher Reflections

In addition to supporting student reflections on knowledge building processes, working with the Think Cards also supported teacher reflections on pedagogy. For example, in the written reflections, one teacher raised his own *I Need to Understand* questions for the other teachers to consider:

*INTU: How do we assess where we as teachers are in terms of the learning of our students? As a community, I feel that we have to move on as a class to create meaning of the lessons and the knowledge that we have built upon. However, how do we determine the cut off time to move on to a different of the lesson if we are constrained in terms of time?*

Further, the teachers got into a discussion about teaching problems that came up in their classes. One example involved students who write "textbook entries" on their cards rather than their own ideas. Another concern involved students who "got all the key science ideas in the first Think Card" and then they didn't know what to do for their New Information card so they wrote copious notes from other resources. These teacher exchanges led to examining more deeply the types of classroom practices and contexts that needed to be created to support knowledge building.

The teachers' discussion of their experiences touches upon the interplay of the nature of disciplinary practices, pedagogy and classroom enactments, and student participation and understanding. We feel that continued engagement in dialogue centered on such interplay is critical in order to support teachers in the creation of classroom knowledge building communities and toward an ever-deepening understanding of what Hogan and Corey (2001) call the "composite culture":

> The "composite culture" represents the classroom culture of science that students actually experience, which is a mixture of ideals of professional science practice… and pedagogical ideals… as filtered through the realities of classroom life and scientific practice. Finally, students' perspectives feed back into shaping the composite culture of the classroom. (p. 216)

This is also consonant with teacher education research on supporting the professional development of beginning elementary science teachers through focused dialogue regarding "(1) engaging in science, (2) organizing instruction, and (3) understanding students" (Mikeska, Anderson, and Schwarz, 2009, p. 678). The teacher discourse indicates that Think Cards can serve as a tool to support dialogue at the intersection of science knowledge building practices, classroom actions, and student understanding.

## Conclusion

The use of Think Cards and hypothetical game-configurations permits guided practice and reflection concurrently with engagement in playing full versions of epistemic games such as the Progressive-Investigation Game. These tools are meant to work in conjunction with Knowledge Forum and video-based reflections in order to help teachers and students to problematise their participation in knowledge building communities. The advantages provided by the Think Cards include physical possession, which creates a strong sense of ownership and the ability to accompany one's ideas into social interactions in order to engage in extended discourse. Students are easily able to exchange ideas and lay out the cards in various configurations in order to see how different ideas are related. One disadvantage over Knowledge Forum is that physical limitations make it difficult for students to see how their ideas fit into the overall structure of the work of the whole class, and students are limited in the ease of creating different "views" on the knowledge base.

Beverly Caswell, a Canadian teacher who worked with the knowledge building communities model in her elementary science classes for many years, describes how as her students learned the processes of knowledge building that "it is almost as if they are thinking 'finally someone is letting us in on the rules of the game of science'" (Caswell & Bielaczyc, 2002). The work described in the present paper concerns ways to put Ideas First classrooms on a trajectory consistent with more skilled and creative playing of the epistemic games of Science. The research contributes to the growing literature on ways to create classroom cultures that mirror disciplinary cultures. Think Cards and hypothetical game-configurations scaffold scientific knowledge building processes, contributing to work with pencil-and-paper based scaffolds (e.g., McNeill & Krajik, 2009) and computer-based scaffolds (e.g., Bell & Linn, 2000). The work presented here captures only a small slice of the Ideas First design and of our work in facilitating a shift from didactic classrooms toward classrooms that function as knowledge building communities. In future work we plan to extend our analysis to classroom discourse around the epistemic game play described here, along with a more in-depth examination of the ways that students use the Think Cards and game-configurations in relation to their work on Knowledge Forum.

## Endnotes

(1) Refer to www.ikit.org
(2) In Singapore, the school year comprises four 10-week sessions. In this school there were 2-2.5 hours of Science scheduled per week in Grades 3 and 4 and an exam period at the end of every 10-week session. There is also a national curriculum specifying science objectives to be covered in preparation for the Primary School Leaving Exam (PSLE), a national high-stakes exam given at the end of Grade 6.
(3) Thinking Phrases are based on the scaffolds found in the Knowledge Forum software. In Ideas First, the Thinking Phrases include: *My Idea is, New Information, A Better Idea is, INTU, Evidence for this Idea, Pulling Our Ideas Together,* and *A Different Idea is.*
*(4)* Scardamalia (2002) outlines 12 design principles for knowledge building communities. In Ideas First we have consolidated these into five key guiding principles. The teacher quotes refer to two of these five principles: *Principle 3: We work together to improve our ideas* and *Principle 5: The community can only move ahead if we help all of the members of our community to move ahead.*

## References


Bell, P. & Linn, M. C. (2000). Scientific arguments as learning artifcats: Designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797-817.

Bielaczyc, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with technology. *Journal of the Learning Sciences*, *15*, 301-329.

Bielaczyc, K. & Collins, A. (1999) Learning communities in classrooms: A reconceptualization of educational practice. In C.M. Reigeluth (Ed.), *Instructional Design Theories and Models, Vol. II.* Mahwah NJ: Lawrence Erlbaum Associates.

Bielaczyc, K. & Collins, A. (2006). Fostering knowledge-creating communities. In A. M. O'Donnell, C. E. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology*. Mahwah NJ:

Lawrence Erlbaum Associates, pp. 37-60.
Bielaczyc, K. & Ow, J. (2007) Shifting the social infrastructure: Investigating transition mechanisms for creating knowledge building communities in classrooms. In *Proceedings of the International Conference for Computers in Education*.
Caswell, B. & Bielaczyc, K. (2002) Knowledge Forum: Altering the relationship between students and scientific knowledge. *Education, Communication and Information*, *3*, 281-305.
Collins, A. (in press). A study of expert theory formation: The role of different model types and domain frameworks. In M.S. Khine and I. Saleh (Eds.). *Models and modeling: Cognitive tools for scientific enquiry*. Dordrecht, The Netherlands: Springer.
Collins, A. & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies for guiding inquiry. *Educational Psychologist*, 28(1), 25-42.
Frederiksen, J. R., Sipusic, M., Sherin, M., and Wolfe, E. (1998). Video portfolio assessment: Creating a framework for viewing the functions of teaching. *Educational Assessment*, 5(4), 225-297.
Goldman, R., Pea, R., Barron, B., and Derry, S. (Eds). (2007). *Video research in the Learning Sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates
Gresalfi, M., & Cobb, P. (2006). Cultivating students' discipline-specific dispositions as a critical goal for pedagogy and equity. *Pedagogies*, 1, 49-58.
Herrenkohl, L.R., Palincsar, A.S., DeWater, L.S., and Kawasaki, K. (1999). Developing scientific communities in classrooms: A sociocognitive approach. *Journal of the Learning Sciences, 8,* 451-493.
Hogan, K., & Corey, C. (2001). Viewing classrooms as cultural contexts for fostering scientific literacy. *Anthropology & Education Quarterly*, 32(2), 214—243.
McNeill, K. L. & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain specific and domain general knowledge in writing arguments to explain phenomena. *The Journal of the Learning Sciences*, 18(3), 416-460.
Mikeska, J., Anderson, A., & Schwarz, C. (2009). Principled reasoning about problems of practice. *Science Education*, 93, 678-686.
Morrison, D. & Collins, A. (1995) Epistemic fluency and constructivist learning environments. *Educational Technology,* 35 (5), 39-45. Reprinted in B. G. Wilson (Ed.) (1996) *Constructivist learning environments: Case Studies in Instructional Design*. Englewood Cliffs NJ: Educational Technology Publications, 107-119.
National Research Council. (2007). Taking Science to School: Learning and Teaching Science in Grades K-8. Committee on Science Learning, Kindergarten Through Eighth Grade. Washington, DC: The National Academies Press.
Ow, J. & Bielaczyc, K. (2007) Epistemic perturbations: Using material artifacts to cultivate a knowledge building culture in classrooms. In *Proceedings of the International Conference for Computer-Supported Collaborative Learning*.
Ow, J. & Bielaczyc, K. (2008). Designing artifacts for "epistemological perturbations" – Changing the bias of learning environments for Knowledge Building. Paper presented at the Eleventh Annual Knowledge Forum Summer Institute, Toronto, CA.
Sawyer, R. K. (2006). Introduction: The new science of learning. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 1-18). NY: Cambridge University Press.
Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.) *Liberal education in the knowledge society*. (pp. 67-98). Chicago: Open Court.
Scardamalia, M. (2004). CSILE/Knowledge Forum®. In *Education and technology: An encyclopedia*. (pp. 183-192) Santa Barbara: ABC-CLIO.
Scardamalia, M. & Bereiter, C. (2006). Knowledge building: Theory, pedagogy and technology. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97-115).Cambridge University Press.
Songer, N.B. (2006) BioKIDS: An Animated Conversation on the Development of Curricular Activity Structures for Inquiry Science. In R. Keith Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences*. New York: Cambridge University Press. P. 355-369.
Thomas, D., & Brown, J. S. (2007). The play of imagination: Extending the literary mind. *Games and Culture*, *2*(2), 149-172.
Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge building communities. *Journal of the Learning Sciences,* 18(1), 7–44.

## Acknowledgements

We would like to thank the Ideas First Team --- both the researchers and teachers who made this work possible. This work was funded by a Singapore MOE grant to the first author through the National Institute of Education's Learning Sciences Laboratory (R8019.735.PM07).

# Multiple Conceptual Coherences in the Speed Tutorial: Micro-processes of Local Stability

Brian W. Frank, University of Maine, 5709 Bennett Hall, Orono, ME 04469-5709, brian.frank@umit.maine.edu

**Abstract:** Researchers working within knowledge-in-pieces traditions have often employed observational approaches to investigate micro-processes of learning. There is growing evidence from this line of work that students' intuitive thinking about physical phenomena is characterized more so by its diversity and flexibility than its uniformity and robustness. This characterization implies that much of the dynamics of students' thinking over short timescales involve processes that stabilize local patterns of thinking, later destabilize them, and allow other patterns to form. This kind of "change" may only involve dynamics by which the system of intuitive knowledge settles into various states without changing the system structure itself. I describe a case study in which a group of college students shift their thinking about motion several times during a collaborative learning activity. Instead of focusing on micro-processes of change, I describe these dynamics in terms of mechanisms that contribute to local stability of students' conceptual coherences.

## Introduction

Conceptual change research in the seventies and eighties largely began focused on characterizing the singular and often flawed ideas that children and students espouse about domain-specific phenomena (Driver & Easley, 1978; Posner, Strike, Hewson, & Gertzog, 1982; Carey, 1986). From this perspective, it was largely assumed that unitary knowledge structures–naïve theories (McCloskey, 1983), alternative frameworks (Clement, 1983), or ontological categories (Chi, 1992)–describe important aspects of novice misunderstandings and that these misunderstanding must give way to other structures in a progression toward expertise. In contrast, more research in the past two decades has focused on characterizing the *diversity* of ideas that children and students employ for making sense of physical phenomena and the nuanced *contextuality* of these varied ideas across time and setting (Strike & Posner, 1992; Smith, diSessa, & Roschelle, 1993; Hammer, Elby, Scherr, & Redish, 2004). From this perspective, it has largely been assumed that a substantial component of expertise involves the gradual refinement and repurposing of existing knowledge–phenomenological primitives (diSessa, 1993), conceptual and epistemological resources (Hammer, 1996), or even pieces of ontological knowledge (Gupta, Hammer, and Redish, accepted).

Both of these perspectives have largely made *change* a central problem to be explained by research. Much of the prior research has focused on broad processes and conditions for change, while the latter has focused on micro-processes of change. In this paper, I focus on describing student understandings as reflecting *multiple local coherences* (Hammer et al, 2004; Rosenberg, Hammer, & Phelan, 2006; Scherr & Hammer, 2008). This specific term is intended to capture the notion that understanding and behavior are often quite variable, while still exhibiting stabilities that are local to moments and settings. While much of this existing work has focused on epistemological coherences, I focus here on describing local coherences that concern the substance of students' thinking. This perspective shifts attention away from seeking only explanations of change and toward explanations of how understandings cohere in specific moments and settings. Locally coherent patterns of student thinking are to be explained in terms of mechanisms that stabilize these patterns. This perspective is explored through an analysis of students' multiple conceptual understandings during a collaborative learning activity in an introductory college physics classroom.

## Interpreting What Distance Means in the Speed Tutorial

Before discussing the theoretical and epistemological considerations for this work, I want to motivate this discussion through an example of student reasoning that is the focus of this study. Two vignettes below are taken from a student discussion during a collaborative learning activity in introductory college physics. The students are in a tutorial—a common instructional format in introductory physics classrooms where students are expected to work together through a conceptual worksheet. In this particular tutorial (a modified version of a tutorial from McDermott, Shaffer, and the Physics Education Group at the University of Washington, 1998), a group of four college students discuss "tickertape" representations of motion.

Tickertape refers to a long strip of paper that can be used to create a record of an object's motion by attaching the paper to the moving object. They are commonly used in physics classrooms to teach students concepts about velocity and acceleration. The record of the motion is produced as the paper is pulled through a tapping device that marks the paper at a constant rate. In this tutorial, strips of tickertape represent the motion of a cart that was recorded prior to class (see Figure 1 for a cartoon schematic). Each student has been given a small strip cut from the entire tickertape that represents the same amount of time (i.e., has the same number of dots) but represents different speeds (i.e., have varying length).

Figure 1: Cartoon Schematic for How Tickertape Strips were Made Prior to Class

Below, the students discuss how the amount of time taken to generate each of their different strips compares, and incorrectly decide (as many groups initially do) that the shorter strips represent less time.

Beth: Obviously, it takes less time to generate the more closely spaced dots.
John: So you are saying it takes less time to make the shorter segments?
Beth: Yeah.
Kate: How can you tell?
Beth: You can tell because it's a shorter distance.
Paul: It's a shorter segment.

The students discuss this conclusion for several minutes, pointing out other features of the strips that support this conclusion. They soon move on to the next part of the worksheet that involves doing some calculations based on distance measurements. After finishing these calculations, the students are prompted by a specific question in their tutorial workbook to reflect upon their assumptions in doing these calculations.

Kate: I think we also assumed in that these were made by the speed at which the paper traveled through the tapper, which was different for each paper.
John: Right, cause if you move it really fast then *[quick movement of hand]*
Beth: That's true! It could depend on how fast the ribbon was pulled *[moves hand]*
Kate: We're assuming that umm-
Beth: That the length is proportional to-
Kate: -The speed at which the ribbon was pulled through.

These two vignettes involve student understandings that are substantively quite different. In the first, students interpret distance as conveying information about time. In the second, students interpret distance to mean speed.

It seems natural, given the obvious change in students' interpretation of what distance means, to ask, "What caused the students to change their thinking?" This focus seems especially relevant since they changed their understanding from an incorrect one (thinking that the distance indicates time) to a correct one (that the distance indicates speed). We might consider several plausible accounts of what caused the change: (i) the particular question from the worksheet triggered them to think meta-cognitively about what they had been doing, (ii) engaging in the measurement exercise enabled the students to attend to and coordinate new aspects, and (iii) Kate was privately thinking the correctly the entire time and only now found a way to have her ideas introduced and considered. Each of these answers may hold some truth in explaining the change in students' collective understanding. The question about change, however, may be misleading.

It turns out that the students don't change their thinking on the matter this one time. Several minutes after the end of the above transcript, Beth brings up again the idea that the shorter strips take less time. Kate, who initially introduced the idea that the speed causes the different distances, seems to now agree with Beth. As the tutorial goes on, the group works to regenerate their understanding that distance must be related to the speed and not the time. However, this is not the last change. As the students go back to a previous page to erase their wrong answers, one of the students is again convinced of their earlier idea. It's not entirely clear that the students ever resolve the issue completely during the instructional period. Instead, their understanding seems to vary between these two distinct ways of making sense of the strips of paper as representations of motion.

For the researcher, trying to understand the cause of just a single moment of change in the students' thinking now becomes a problem of explaining multiple changes back and forth. The seemingly straightforward question, "What caused the students to change their thinking?" may not only be difficult to answer, it may be misleading in its premise. For starters, there might not be a simple or single causal explanation for any of these transitions. In addition, the apparent change in the students' thinking might not reflect change to the structure of students' intuitive knowledge at all. Instead, the changes we observe might only reflect dynamic transitions among competing understandings that each exhibits some local stability. Understanding the local stabilities themselves becomes a central problem for researchers to address. The aim of this paper is to describe how the students' initial understanding may be stabilized through a variety of different mechanisms.

## Theoretical Orientations

The shift in perspective I am advocating coincides with an underlying assumptions that distinct patterns in student thinking reflect dynamic stabilities and that these stabilities require explanation. The stability of student thinking need not be due to the existence of a robust belief or mental category, but rather due to how real-time activity creates self-sustaining patterns of thought. The notion of real-time activity has often been discussed dichotomously with respect to human behavior and learning—either encompassing individual, cognitive activity (e.g., connectionism) or as distributed, social activity (e.g., activity theory). In this section, I briefly discuss relevant aspects of these individual and activity perspectives.

### Activity among Knowledge Agents

From individual cognitive perspectives, complex knowledge systems approaches describe activity as occurring among fine-grained knowledge structures. Minsky (1986), for example, conceptualized the individual mind as a society of mindless "agents"—numerous in kind and in their interactions. Agents, like schemata more generally conceived (Bartlett, 1932/1995; Rumelhardt, 1980), can be thought of as cognitive elements for perceiving, doing, and remembering. These pieces come together in various ways to generate and represent thinking. One kind of knowledge piece that has been discussed quite extensively is diSessa's (1993) phenomenological primitives—bits of knowledge reflecting our intuitive sense of mechanism. In the above transcript, we might think of the students' inference about the shorter strips taking less time as resulting from the intuition that *less distance implies less time*. This bit of knowledge for relating space and time is part of how the students interpret and make sense of the strips as representations of motion.

Beyond just describing kinds of knowledge, diSessa distinguishes between two properties of intuitive knowledge systems that he calls *cuing priorities* and *reliability priorities*. These two properties drive mechanisms that activate and stabilize ideas in particular contexts. Phenomenologically, cuing priorities describe the likelihood that a given piece of knowledge will activate in a given context, while reliability priorities describe the likelihood that a given piece will remain active in a given context. In the above example, we see the idea "shorter strips take less time" becoming focal in the conversation and then persisting for several minutes beyond the transcript.

Structurally, if one thinks of knowledge elements in terms of a network, reliability is related to the number of activation pathways leading away from and then back to a given knowledge element. In other words, a given idea can be made more stably active (in a given setting) because other "nearby" ideas support its sustained activation. Thus, the relationships and interactions among many ideas serve as one kind of mechanisms by which patterns of thinking are self-sustaining over a period of time. With the transcripts above, we whether other ideas arise during their conversation that help to support their continued thinking that the shorter strips take less time, as well as how the context supports the cuing of those ideas.

### Activity among Participatory Agents

From a more distributed and social perspective, activity-theoretical and other situated approaches describe activity as occurring among persons and artifacts, not among entities of the mind. In many respects, this view represents an oppositional stance to the above knowledge-based accounts. Lave (1988), for example, in describing the variety of mathematical practices, denies the existence of mental constructs such as arithmetic. Despite this dichotomy, there are those who do attempt to capture both knowledge and participation views in their accounts of students' thinking (e.g., Roschelle, 1998) or have advocated for their commensurability.

One major difference between knowledge- and participation-based accounts concerns the typical unit of analysis. The unit of analysis for situated perspectives is often described as the entirety of *persons-in-settings* (rather than individuals or individuals' knowledge). It is more focused on observable unfolding social activities (than knowledge use) and how these activities arise within social settings that are defined by the activity (and not by platonic outsider descriptions). Beyond just focusing on humans, these perspectives focus on describing how material artifacts mediate aspects of that activity (Engestrom, 1987), provide affordances for action (Greeno, 1998), and shape semiotic fields (Goodwin, 2000). In the vignettes above, the students do not only have different understanding. They also interact with each other and their settings in quite different ways.

Different patterns of action and social collaboration, taking place within changing material settings, can be understood to constrain and support activities that also differ in terms of the *substance of the ideas* to which activity is oriented. Embodied cognitive perspectives, for example, emphasize how human understandings have a strong basis in bodily action (e.g., Thelen & Smith, 1994). Pointing is an action that might support ideas for understanding objects and their spatial relations, while dynamically moving one's hand through space might better support ideas for understanding trajectory and cause and effect relations. Similarly for social collaboration, establishing joint attention as part of collaborative action might afford different epistemic activities than establishing mutual attention. Such social cues help individuals to *frame* (Scherr & Hammer, 2009) activities.

Taken together, the affordances of artifacts and the patterns of participation that take shape around them serve as mechanisms by which activities stabilize in settings. By carefully considering the substance of ideas to which these activities orient as well, the analytical task undertaken in the next section is to explain the stability of students' understanding in terms of both knowledge- and participation-based mechanisms.

## Explaining the Local Stability of Reasoning

The goal of this section is to develop an explanatory account of the stability of the students' initial thinking in terms of real-time processes. I describe several knowledge- and participation-based mechanisms that plausibly contribute to the stability of students' thinking and connect these claims to specific evidence.

### Knowledge Based Mechanisms

#### Contextual Feedback from Setting

One way in which the students' initial thinking that "shorter strips take less time" stabilizes is due to how particular micro-features of the context support the continued activation of underlying intuitive knowledge. Several aspects of the particular context in which students discuss their ideas support the persistence of the idea that *less distance implies less time*: the specific language of the worksheet question, the salience of length differences among the tickertape strips, and a congruence of part-whole relationships on the strips.

The particular question from the worksheet reads, "How does the time taken to generate one of the short segments compare to the time to generate one of the long ones?" Beth immediately states, "Obviously, it takes less time to generate the more closely spaced dots." The particular phrasing of the question not only draws attention to the distance features of the strips, it uses the words "shorter" and "longer" to do so. These specific words are relevant, because they are used flexibly in everyday language to both refer to distance and time (but no other concepts). The ambiguity of these words closely ties with the intuition that *less distance* ("shorter") *implies less time* ("shorter"). The students repeatedly use those words while discussing their initial ideas.

The students' understanding is also supported by features of the tickertape strips. By far, the most salient feature of the strips is that they are different lengths. One can observe this when standing far away, without closely inspecting the details of where the dots are located. The salience of this feature supports students' initial and sustained attention to distance features of the strips, which has been shown to preferentially cue the idea that *less distance implies less time* over other competing ideas when students are asked to made judgments about duration (Frank, Kanim, Gomez, 2009). In the transcript, after Beth initially gives her answer, the students explain that they can tell its less time because the strips are shorter. As they explain this, the students can be seen reaching toward the strips and indicating distances on the strips with their fingers and pencils. These statements and gestures show that students are closely attending to this salient feature in support of their initial understanding.

A second feature of the strips that contributes some stability arises from the fact that each of the strips has the same number of dots. When inspecting the strips, students do not only attend to the total length of the strips, but also to the length of the spaces between dots. Because each strip has the same dots, this means that the longer strips are made up of longer pieces, and the shorter strips are made up of shorter pieces. As students shift their attention from the entire length to the length of the parts (or vice versa), students see the same information. Shorter strips or shorter parts imply a shorter amount of time. Evidence that this shift in attention does not disrupt the students' thinking, and may in fact stabilize it, comes from a part of their conversation that happens shortly after the end of the first transcript above.

Kate: When we are talking about segments, are we like not thinking about how long the total paper is? Are we just looking at the marks. Are we supposed to be considering
John: I'm guessing they like mean from here to here. *[pointing with pencil between marks]*
Kate: Like I wonder why like the papers are all different lengths.
Beth: Cause none of these papers are the exact same si-ize. Except for these two *[pointing]*
Paul: Right because I think *[moving hand toward the center]* they all have the same amount of dots *[pointing to several locations on one of the strips]*.
Beth: Oh-oh
Paul: I think they all have six dots.
Beth: Oh do, they?
Kate: Is that true? 1, 2, 3, 4, 5, 6 *[pointing to successive dots on a strip]*
Paul: So, it's a shorter amount of time for a shorter piece of paper

Kate asks if they are supposed to be looking at the distance between the "marks" or the total length. As the students coordinate between these two features (realizing that there are six dots on each), the students maintain their understanding that the shorter strips take less time.

### Structural Feedback from Other Knowledge

The students' initial understanding is also stabilized through the recruitment of other ideas that support its persistence. Two specific ideas arise together with the idea that the shorter strips take less time. One is that *bunched up means faster,* and the other is *faster implies less time.*

Students attend to the length of the entire strip and to the length of its parts, but they also attend to the "density' or overall proximity of the dots to each other. These students, as well as many others, describe the bunched up patterns as "faster" and the more spread out patterns as "slower". The words "faster" and "slower" used here is another example of ambiguous and flexible use word meaning, similar to how "longer" and "shorter" are used in reference to the distance and time. The students notice "more dots happening in less space" and describe this pattern as "faster". However, the same word is used by these students to describe a greater speed and also a lesser the amount of time, and therefore provides some ambiguity about its meaning.

Both immediately before and after the students discuss that the shorter strips takes less time, several of the students refer to the shorter strips (with the more closely spaced dots) as faster:

Beth: Nobody has the exact same rate
John: *[pointing to the shortest strip]* So I guess that's the fastest.
… (and then later in their discussion)…
Beth: How do you know how to arrange –
Kate: Ahh. The shorter the segments the faster the speed
John: Yeah.
Paul: Also shorter the paper, it's the same thing.

Here we see the students using the word faster to identify shorter segments. The idea *bunched up means fast* can be understood to work together with another idea that arises. In this case study, and many others described elsewhere (Frank, 2009), we see evidence that students discuss the idea that *faster implies less time.* Beth explicitly discusses this idea during their discussion at a slightly later time when she says, "Wouldn't yours be going slower than mine, because it took more time to make that same?" Her statement implies that a shorter strip goes slower because it takes less time, which she and the other students had established earlier as being known because the strip was shorter (bunched up). A single strip is thus conceived of as being short, bunched, and fast, with the words "short" and "fast" having multiple meanings that overlap.

These three elements of intuitive knowledge– *less distance implies less time, bunched up means faster*, and *faster implies less time*– are all connected through their linguistic and conceptual overlap. Ideas of *less distance* and *less time* are bound by a linguistic overlap with the words "longer" and "shorter". *Less distance* and *bunched* are bound by their conceptual overlap with a spatial sense of proximity. *Bunched* and *less time* (and even notions of speed) are bound by their linguistic overlap with the word "faster" and "slower". This network of ideas can be understood to exhibit stability through the mutual relationships among its parts.

These knowledge-based mechanisms describe how relationships among specific elements of knowledge and contextual features provide some stability to their understanding, contributing to its persistence over time. In the following section, I describe aspects of activity that provide stability in different ways.

## Participation Based Mechanisms

The previous section largely focused on mechanisms that concern how knowledge elements remains active due to interactions with features of context and with other elements of knowledge. In this section, I discuss how the students' thinking is stabilized through processes by which students interact with each other and with artifacts.

### Feedback among Interactional Behaviors

Students' collective interactions with each other and with various artifacts around them show distinct patterns that are relevant to understanding their activity. For the first fifteen minutes, the students collectively move in and out two broad clusters of interactive behavior. These behaviors largely occur together among all the individuals, with some exceptions. The two distinct patterns are (i) behaviors oriented toward their worksheets (see Figure 2)- a pattern described by Scherr and Hammer (2009)- and (ii) behaviors oriented toward the strips at the center of the table (see Figure 3).

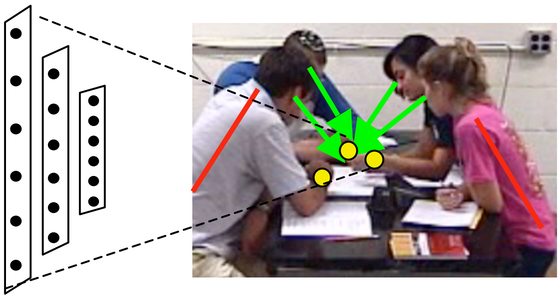


Figure 2. Student Behaviors Oriented toward the Strips

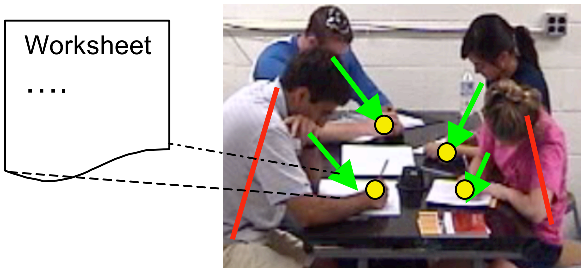



Figure 3. Student Behavior Oriented toward Worksheets

These behaviors support important aspects of the students' activity, which establish and stabilize a setting for their discussion. The tickertape-oriented behaviors are characterized by patterns of interaction that involve directing gaze, gesture, and bodies toward the strips at the center of the table. Students lean in, look inward, and position their hands at the center of the table. The students' clustering in this behavior is dynamically stable in the sense that any one student's behavior is coupled to the behaviors of the others. As students lean toward the center and point to strips and locations, other students look in that direction, and mirror these actions by leaning in and pointing to features as well. While individual students opt in and out, the collective structure of participation persists over much longer time scales. The worksheet-oriented behaviors are characterized by different patterns of interaction: hunching over, looking down, and using hands to write and point to worksheet locations. This pattern also exhibits a dynamic stability.

What is striking about the students' behavior during this time is the remarkable synchrony with which they move in and out of these two patterns. It is a highly coordinated (but not centrally directed) activity of noticing and describing patterns and then writing down answers to specific questions. This activity of establishing joint attention to community objects may be understood to couple to the substance of their ideas as well. Their initial understanding during this time concerns inferences drawn from noticing and comparing features of the strips (e.g., *short distance means short times* and *bunched up dots means fast*). These externalized behaviors for establishing joint attention are part of the dynamic by which the students access ideas and employ strategies used for noticing, describing, and comparing patterns across the strips.

Importantly, during this time, the students' neither discuss aspects of the physical motion of the strips, nor do they engage in any significant moments of mutual attention, as they do later. The above behaviors contrast with their later behaviors, where they orient their behaviors to each other—sitting up right, looking to each other, and gesturing in animated ways.

## Material Affordances and Constraints

The two patterns of collective behaviors above, which help to characterize the students' initial activity, take place in a particular material setting that provides affordances for these behaviors. The students themselves play a role in constructing this setting. In particular, during the time that the students' initially persist in thinking that the shorter strips take less time, the strips are located and arranged at the center of the table. The students moved the strips to this location and arranged them in order by length at the very onset of the tutorial. This centralized location allows the students to collectively orient to the strips and sustain an activity of pointing out and describing features of the strips.

The arrangement in the side-by-side fashion serves two purposes. First, it offloads much of the task of noticing and remembering to the setting. Second, the arrangement has affordances for particular kinds of gesture. Students gesture in various ways (pointing with a finger, indicating a space with two fingers, etc), but all of the gestures are deictic. These embodied actions may support students' ideas for thinking about objects and spatial relations among parts (the density of dot, the number of dots, etc.)

The anchoring of their activity to the strips as material artifacts and the dynamic stability of their collective behaviors feed into many of the knowledge-based mechanisms described in the previous section. Their individual behaviors couple to each other and to material objects. These behaviors help to sustain particular patterns of attention and action upon the strips. These behaviors support their use of specific intuitive knowledge. In one sense, we can think of the stability of students' thinking as coupled to a physical stability of the world. The strips, once placed at the center of the table, remain put until some actively moves them again. Similarly, the students have located their worksheets so that they are closely hugged to their bodies, making it easy to coordinate their dual activity between noticing and describing, and filling out their worksheets. The students seamlessly move back and forth between these two patterns of behavior repeatedly during the first ten minutes of tutorial. Their careful attention to just the visual properties of the strips (i.e., not to any sense of motion) in order to answer specific worksheet questions takes place in a dynamic setting that supports its persistence.

Just as students' thinking and behavior during the first episode contrasted with their later thinking and behaviors, the students' material setting later is different as well. Their later understanding happens in a setting where the students have separated and decentralize the strips. This change to their material settings provides new opportunities for embodied action and participation that did not occur before. In the second vignette, where Kate explains how the strips were made, Kate can be seen lifting a strip off the table and physically enacting the pulling of the strips by the cart. In doing so, the other students in the group look to Kate. The students engage in a new pattern of mutual attention as they discuss aspects of the physical motion which made the strips. This is the first time they engage in significant mutual attention, and is part of the dynamic by which they in articulate and listen to more complex ideas.

## Discussion

The primary goal of this paper is to characterize the stability of a single group of students' initial thinking during a collaborative learning activity in terms of a variety of micro-processes that together help to sustain one possible understanding for several minutes. A full analysis of these dynamics would also need to include an account of the mechanisms that contribute to their later thinking as well, but is beyond the scope of this paper. Given the dynamic changes apparent in their understanding during the tutorial, I claim that it is productive to focus on explaining the local persistence of understanding, rather than to begin with focus on causes of change.

One distinguishing feature of this analysis is its focus on explaining this stability both in terms of how elements of students' intuitive knowledge are reliably activated in the setting and in terms of how students' interactional behaviors dynamically constitute and stabilize aspects of this setting. I claim that the students make their initial inference about the tickertapes based on a shared intuition that *less distance implies less time*. There is evidence that each of the students makes sense of the situation using this idea at various times, and often times they do so together. I describe how particular features of the context (as characterized by the existence of specific features on worksheet question and on the strips themselves) contribute to activation of that knowledge, and also describe how other knowledge recruited to make sense of the situation provide a stable "network" of ideas for making sense of the patterns they notice. This network of ideas, however, arises within a social context of discussing patterns and writing down answers to worksheet questions. During their initial discussion, the students orient their gazes, gestures, and bodies toward the collection of the strips at the center of the table and toward the worksheets in front of them; but they do not engage in behaviors of mutual attention until much later. This activity of noticing and describing patterns to be written in their worksheets is both stabilized by the students' specific behaviors in this activity (to which they collectively orient) and by the location and arrangement of material artifacts that are central to the activity.

This characterization, while focused on explaining the stability of the students' initial pattern of thinking, does provide insight into the dynamics of their subsequent understanding as well. Instead of seeking explanations in terms of "causes" for the change, we can, instead, see how the various constraints imposed by their activity change. These changes in the constraints of their activity provide opportunities for new patterns to take hold with their own local stability. The students' new thinking co-evolves with changes to their patterns of attention (now to physical mechanisms involved in the motion), changes to the location of material objects (the strips become separated and within several of the students hands off the table), and to the patterns of interactional behavior (now they look at each other as they articulate complex ideas). Just as there may be not single cause for the local stability of the students' thinking, their may be not single explanation for why the students' thinking changes either. Across the entire tutorial, there are many changes that are more-or-less continuously changing over time. Coming to understand which aspects of activity are merely ephemeral and which continue to exert influences is a major goal of further research in this direction.

Beyond this particular case, considering students' intuitive thinking as reflecting multiple conceptual coherences may be useful for describing stabilities of thinking and behavior that seem to persist on longer time scales (e.g., Why do students appear to hold on to robust misconceptions?). It may be possible to both account for the variability of student understandings while also describing how, in particular contexts, it can settle into patterns indicative of common and robust misconceptions. This program might be achieved through careful attention to the multitude of cognitive and social mechanisms that contribute to the local stability of reasoning as it occurs within activity.

## References

Bartlett, F. C. (1932/1995). Remembering; a study in experimental and social psychology. Cambridge [Eng.]: The University press

Carey, S. (1986). Cognitive science and Science Education. *American Psychologist, 41*(10), 1123-1130.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. Giere (Ed.), *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science* (pp. 129-186). Minneapolis, MN: University of Minnesota Press.

diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction, 10*(2/3), 105-225.

Driver, R., & Easley, J. (1978). Pupils and Paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education, 5,* 61-84.

Engestrom, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research.* Helsinki: Orienta-Konsultit.

Frank, B.W (2009). *The dynamics of variability in introductory physics students' thinking: examples from kinematics.* Unpublished Doctoral Dissertation. University of Maryland, College Park, MD

Frank, B.W., Kanim, S.E., & Gomez, L.S. (2008). Accounting for variability in student responses to motion questions. *Physical Review Special Topics - Physics Education Research, 4,* 020102

Greeno, J. G. (1989). A perspective on thinking. *American Psychologist, 44*(2), 134-141.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*(10), 1489-1522.

Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics, 64*(10), 1316-1325.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. (2004). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of Learning: Research and Perspectives.* Greenwich, CT: Information Age Publishing.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life.* Cambridge, England: Cambridge University Press.

McCloskey, M. (1983). Naive Theories of Motion. In D. Gentner & A. Stevens (Eds.), *Mental Models* (pp. 299-324). Hillsdale, NJ: Erlbaum.

McDermott, L. C., Shaffer, P. S., & the Physics Education Group at the University of Washington (1998). *Tutorials in Introductory Physics.* Upper Saddle River, NJ: Prentice-Hall.

Minsky, M. L. (1986). *Society of Mind.* New York: Simon and Schuster.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accomodation of a scientific conception Toward a theory of conceptual change. *Science Education, 66*, 211-227.

Roschelle, J. (1998). Beyond romantic versus sceptic: a microanalysis of conceptual change in kinematics.. *International Journal of Science Education, 20*(9), 1025-1042.

Rosenberg, S. A., Hammer, D., & Phelan, J. (2006). Multiple epistemological coherences in an eighth-grade discussion of the rock cycle. *Journal of the Learning Sciences, 15*(2), 261-292.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. Bruce & W. F. Brewer (Eds.), Theoretical Issues in Reading and Comprehension (pp. 38-58). Hillsdale, NJ: Erlbaum.

Scherr, R.E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction, 27*(2), 147 – 174.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *Journal of the Learning Sciences, 3*(2), 115 - 163.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice* (pp. 147-176). Albany, NY: SUNY.

Thelen, E., & Smith, L. B. (1993). *A Dynamic Systems Approach to Development.* Cambridge, MA: MIT Press.

Vosniadou, S., & Brewer, W. F. (1987). Theories of Knowledge Restructuring in Development. *Review of Educational Research, 57*(1), 51-67.

## Acknowledgments

The research described here was conducted at the University of Maryland. I'd like to thank Rachel Scherr, David Hammer, Joe Redish, Andy Elby, Ayush Gupta, Luke Conlin, Renee Michelle Goertzen, and Michael Wittmann for their helpful conversations. The research has been funded in part by the National Science Foundation under Grant Nos. REC-0440113 and REC-0633951. Any opinions, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

# Gaining an Insider Perspective on Learning Physics in Hong Kong

Jan van Aalst, The University of Hong Kong, Hong Kong, SAR, China, vanaalst@hku.hk

**Abstract:** This study provides an exploration of physics teaching and learning in a classroom in Hong Kong. The goal of the study was to understand how to develop Bereiter and Scardamalia's knowledge building approach for Asian contexts in away that is sensitive to Asian values. The paper reviews some aspects of learning physics in Asia from a Confucian-heritage perspective, and then reports a case study of physics learning with Knowledge Forum. Participants were 82 Form 4 (Grade 10) students taking two successive versions of a physics course. Data collected included classroom observations, measurement of physics learning, reflection, epistemic beliefs, attitudes toward science, and use of Knowledge Forum. The findings open up a number of questions for further research.

## Introduction

Work in classrooms based on the major educational approaches produced by the learning sciences is still at an early stage in Asian contexts, particularly approaches that make extensive use of collaboration, inquiry, and classroom discussions. However, to advance from this position does not merely involve *adapting* Western approaches to Asian contexts, which may lead to misinterpretations of how learning occurs there (Biggs, 1996). Rather, the premise of this paper is that it would be useful to develop an insider perspective on the learning sciences, based on a study of the values underpinning educational practices, review of research on Asian classrooms, extensive classroom observation, and a critical examination of perspectives from the learning sciences. With an insider perspective, we may then develop educational approaches in ways that build on Asian values and accomplishments, without inflicting what Ann Brown called "lethal mutations" (Brown & Campione, 1996).

This paper reports on a two-year exploration of this kind focusing on the teaching and learning in two successive Physics courses (Form 4 or Grade 10) taught by the same teacher; the context for this exploration was a research program on knowledge building (Scardamalia & Bereiter, 2006). The researcher reviewed the literature on cultural influences on science education in Confucian-heritage countries (mainly China, Korea, Japan, Taiwan, Singapore, and Malaysia), used ethnographic methods to observe classes, and measured wide range of variables thought to be relevant to knowledge building. Assessments measured physics knowledge, reflection on knowledge, attitudes towards science, and epistemic beliefs. Assessments of physics knowledge included capabilities considered important by the Physics Education Research (PER) community—detailed knowledge of concepts, and ability to develop explanations that refer to physics principles and causal mechanisms. The goal of the paper was to *benchmark* performance by the classes studied against international studies of physics learning and knowledge building, and to gain an understanding of how to evolve the knowledge building approach. Suggestions for development are outlined in the conclusion section.

## Cultural Considerations

External examinations are important in the West and influence teaching, but they have far greater cultural significance in the East. From the Han Dynasty (206 BC-220 AD) to the end of Imperial China, the *Keju*, a set of government examinations was used to select people from all walks of life for government positions (Amano, 1990). Preparing oneself for these examinations involved years of effort and commitment, but when a person passed them it led to upward social mobility and was a matter of great pride and advantage for their family. As an ancient Chinese idiom states, "Although studying anonymously for 10 years, once you are successful, you will be well-known in the world." The influence of Confucianism has decreased throughout the 20th century, but the emphasis on effort and accomplishment and their relationship to honor and social mobility have endured, producing societies that remain more hierarchical than in the West. Scholars have observed that moral self-perfection and social development are the most important educational goals, ahead of intellectual achievement (Gao, 1998; Lee, 1996; Li, 2009). Lau and colleagues concluded from a recent study of management issues in science classes in China, Israel, and Australia that Chinese teachers were more likely to mention that "learning to respect authority was a significant outcome of education" (see Lewis, Romi, Qui, & Katz, 2005, p. 731).

In present-day Hong Kong, students present themselves for the Hong Kong Certificate of Education Examination (HKCEE) at the end of Form 5 (Grade 11). Consistent with the above cultural influences, these examinations have a high standard. For example, of 29,713 students taking the 2008 Physics examination, only 4.6% received an 'A', 30.4% between 'A' and 'C', and 77.6% any passing grade (Hong Kong Examination and Assessment Authority). HKCEE results are very important for social mobility. Many employers require at least five passes for clerical positions, and HKCEE results are taken into consideration, besides university results, in applications to postgraduate programs. Standards in other Asian jurisdictions are similar. In a study of Chinese teachers (Gao & Watkins, 2002), one teacher commented that performance on matriculation examinations is "the most important even the only aspect by which the school assesses my teaching" (p. 65). These authors assert: "If the performance of students is not as good as expected, their teachers, principals and the head of the local government education department are all punished" (p. 71). Clearly, in implementing innovative educational approaches, teachers cannot afford to jeopardize examination results. Although in Hong Kong substantial curriculum and assessment reforms are in progress, there is no evidence that the high standards will be adjusted.

These conditions have a profound impact on classrooms. For example, Gao (1998) found that in China many teachers teach a three-year physics course in two years, reserving the final year for review and practice for the examination. Teachers may also schedule extra lessons and homework during school vacations. These practices are used because although daily teaching focuses on understanding of physics, this is not thought to be sufficient for the level of precision and detail at which students are examined. Asian teachers prefer didactic approaches, but disagree these are teacher-centered. One Chinese teacher interviewed by Gao and Watkins (2002) said, "If the teacher focuses on encouraging students, setting questions to challenge them, directing them to explore new knowledge, I don't think that means teacher-centered" (p. 73). Chinese teachers also say that classroom learning is not the whole of student learning, that after class student learning becomes more active (Gao, 1998). Several studies conclude that Taiwanese students have few opportunities to discuss their ideas in class (Aldridge & Fraser, 2000; Wallace & Chou, 2001). Aldridge and Fraser refer to a 1984 study on child rearing in Hong Kong, which concluded that attitudes towards filial piety (respects for one's parents and ancestors) was correlated positively with strictness and discipline and negatively with the expression of opinions and independence (p. 214). According to some scholars, Chinese students prefer to learn (comprehend) new material before raising questions about it and discussing it with others (Li, 2009).

Nevertheless, pedagogical approaches have begun to involve more activities and discussions. In a study of Taiwanese and Australian students, Aldridge and Fraser (2000) found that Taiwanese students focused more on the learning task, and the Australian students on the social interactions; although the Taiwanese students perceived their learning environment as more task oriented, they rated it more positively than the Australian students. Wu and Huang (2007) concluded that Taiwanese low-achieving ninth grade students in an interactive environment "did not receive direct support from the teacher that could constantly draw their attention to the content, [and] had fewer opportunities to listen to or engage in thoughtful discussions about concepts" (p. 747). These studies suggest that in Asian contexts it would be difficult to implement a constructivist approach in a Western sense, and that substantial teacher guidance is needed in addition to activities.

It is important to note that international studies of achievement in science such as the TIMSS and PISA consistently place Asian countries well above the international average (Leung, Yung, & Tso, 2005). Clearly, innovative educational approaches should not jeopardize such positive outcomes. Whereas Western observers have considered learning in Chinese classrooms as rote and the positive outcomes as "paradoxical" (Biggs, 1996), scholars now believe that a memorization-understanding strategy proposed by Ferrence Marton can account for the paradox (Chan & Rao, 2009). Preparation for Chinese examinations requires a high degree of practice and memorization, which is facilitated by conceptualization. However, although the results are strong in terms of subject knowledge, the various strategies used to prepare students for examinations may contribute to "learned helplessness" and keep students reliant on teachers. This is why *learning how to learn* has become an important emphasis in curriculum reforms in Hong Kong (EMB, 2000). At the same time, in the 21st century, people need to be able to collaborate and deal with novel situations. Perhaps the learning sciences can make a contribution to innovate learning approaches in Asian schools by developing new approaches to the memorization-understanding strategy that support the development of higher-order thinking and collaborative skills.

## Methods

### Participants

Participants were 82 students from two successive cohorts of a Form 4 (Grade 10) physics course, taught by the same teacher at an English Medium of Instruction (EMI) school in Hong Kong. The school was classified as medium academic achievement based on previous HKCEE results. Lessons were conducted in Cantonese, but all learning materials were in English, and students completed all assignments and tests in English.

### Goals and Design

The study is based on the first year of a two-year physics course; topics taught in the first year included heat and temperature (3 months), mechanics (5.5 months), and waves (1.5 months). The goal of the teacher's work was to explore how to improve on her previous efforts to promote knowledge building (Scardamalia, 2002; Scardamalia & Bereiter, 2006), an approach that involves collaborative inquiry and discourse aimed at advancing collective knowledge in a community. The use of Knowledge Forum®, a web-based inquiry environment, is an important aspect of this. Specifically, the teacher aimed to develop more collaborative discourse in both the classroom and Knowledge Forum, and to improve the integration of the use of Knowledge Forum with classroom learning.

In the first cohort (Class A, 2007 to 2008), students used Knowledge Forum to discuss their own questions about physics content (e.g., a discussion to examine why Eskimos live in igloos, which would intuitively seem cold dwellings). They also used it for discussions about two short projects, in which they collaborated in small groups: designing a solar cooker, and studying motion of their own bodies on a ride at a local amusement park. Both of these projects were heavily guided by the teacher. In the second cohort (Class B, 2008 to 2009) some modifications were made to this design: An interactive white board (IWB) was installed to facilitate easier crossing between classroom talk and work in Knowledge Forum, and the amusement park projects were replaced by projects in which groups of students designed their own experiments to investigate motion through a resistive medium. However, the teacher's workload was heavier that year, and Knowledge Forum was started three months later than in the previous year.

## Data Collection and Analysis

### Classroom Observations

To gain an understanding of the classroom ethos and discourse, the researcher observed double lessons conducted in the physics lab once to twice per month throughout the project; approximately 400 photographs and 20 hours of video were collected. To facilitate discussions with the teacher, several short movies (5-7 minutes each) were produced from these materials. Choices about what to emphasize in these reified the researcher's interpretation of what was happening that was relevant to knowledge building. These movies also were discussed with several other teachers and researchers working on knowledge building in South-East Asia. Interpretations were also revised based on extensive reading of the literature on teaching and learning in Confucian-heritage contexts (summarized earlier). Thus, interpretations were triangulated in several ways.

### Assessments of knowledge, learning process, epistemic beliefs, and attitudes

Extensive data were collected to measure what students were learning and to explore its relationship to work on Knowledge Forum, epistemic beliefs, and attitudes towards science. As much as possible, instruments were used that have been used widely in Western contexts, and that were validated for Hong Kong students. Some additional assessments consisted of assignments designed by the teacher or researcher. Due to space limitations only results for the mechanics unit are discussed here, see Table 1 in the results section.

*Force Concept Inventory (FCI):* This survey remains very popular in the PER community and has been administered to thousands of students (Hake, 1998; Hestenes, Wells, & Swackhamer, 1992). Several items that related to circular motion and projectile motion (which were not part of the curriculum) were removed.

*Learning reflections:* At the end of the mechanics unit students were asked to write a short essays to respond to two questions: "What is your big learning about mechanics?" and "What is the most difficult question you now have about mechanics, and what would you need to know to understand it?" Most students wrote up to one page for each question. Protocols from the heat and temperature unit from Class A were used to develop three scales with acceptable inter-coder reliabilities: *Content Knowledge*, *Everyday Applications*, and *Self-Awareness*. Content Knowledge refers to students' ability to recall the main points of what they had learned about physics concepts and phenomena; at the highest level of the 1 to 4 scale students needed to reveal insight into this knowledge (e.g., point out inter-relationships between ideas). Content Knowledge is a general indication of physics knowledge, whereas the FCI provides detailed information about students' conceptual understanding. Daily Applications similarly assessed students' ability to use their knowledge of physics to explain phenomena from daily life. This ability is needed for the HKCEE, in which students are required to answer questions about a short newspaper article that involves physics knowledge. Self-awareness is a measure of students' ability to identify the strength and weaknesses of their knowledge and how they were able to learn. For simplicity of presentation, we converted the results to percentages.

*In-class tasks*: Two tasks were designed to assess students' understanding and explanations during the unit. A *Graphing Task* was given after the first lesson on motion graphs; students were asked to draw *s-t*, *v-t*, and *a-t* graphs relating to riding a bicycle around the school, and to state their main idea, something they wondered about, and something they wanted to raise for discussion. At the end of the mechanics unit, students used a cartoon involving a horse and cart, in which the horse argues that it is impossible to get the cart moving, and were asked to state, with reasons, whether they agreed with the horse (Explanation Task, N3).

*Epistemic beliefs*: It is assumed that the knowledge building can lead to more sophisticated views about knowing and knowledge. Therefore, epistemic beliefs were measured using a Chinese version of the Epistemic Belief Inventory (28 items on 7-point Likert scale, see Schraw, Bendixen, & Dunkle, 2002). The questionnaire was validated using 484 Hong Kong students taking Form 3 and Form 4 science courses. Exploratory and confirmatory factor analysis revealed three factors with acceptable reliability: *Innate Ability*, *Simple Knowledge*, and *Quick Learning* (alphas 0.75, 0.68, and 0.66). Li (2009) suggests that Chinese learners do not hold Native Ability beliefs but the relatively clear factor structure suggests that Hong Kong science students *do* hold them. Preliminary analysis showed that there were no statistical differences between pre-test (start of the school year) and post-test (end) results; therefore, only the post-test data are reported.

*Attitudes Toward Science*: A 14-item questionnaire (5-point Likert scale) was used to measure attitudes towards science such as anxiety and interest at the end of the year. It was validated using 101 Form 4 science students taking science (Chronbach alpha 0.93).

*Contributions to Knowledge Forum*: Six indicators of contributions to Knowledge Forum were measured (see Table 1 for details). These have been used widely in previous studies involving Knowledge Forum.

## Results and Discussion

### Classroom ethos and discourse

The four pictures in Figure 1 are intended to illustrate the learning environment that was observed and then enhanced by introducing an IWB.

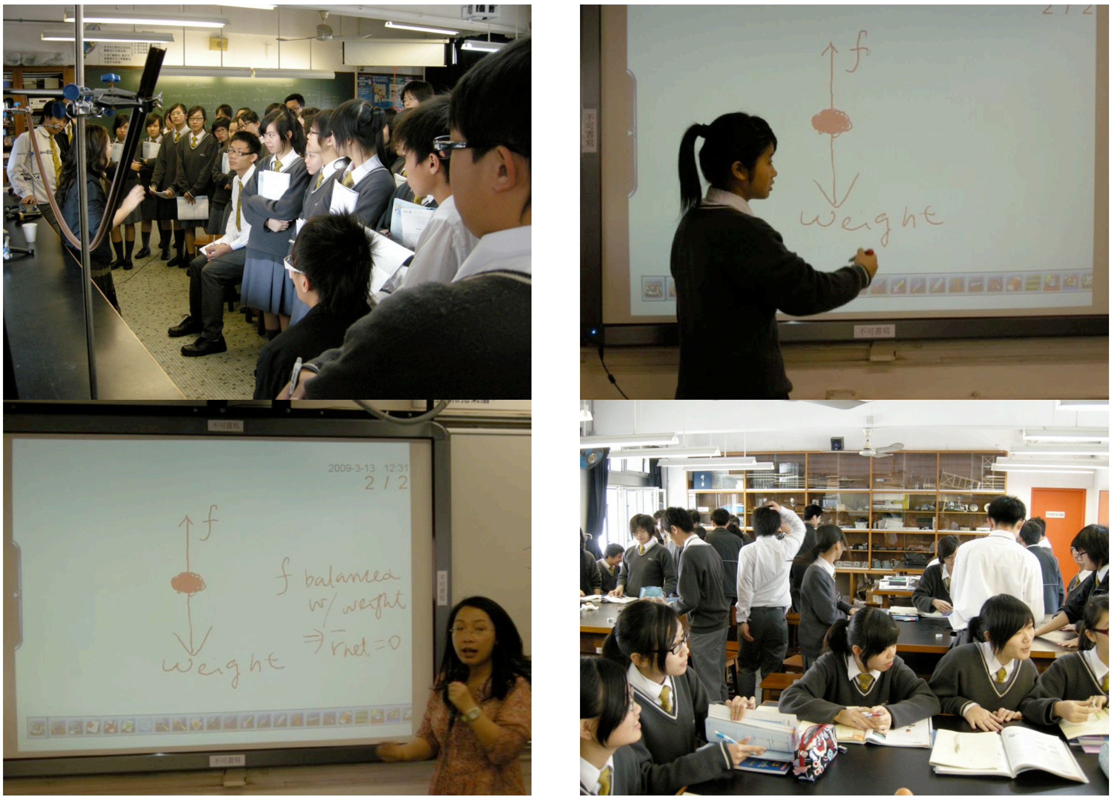



Figure 1: Classroom Environment

At the *top left*, the students gather around the teacher, who is demonstrating an experiment involving conservation of energy; the concept has already been studied, but this time there is not enough time or equipment for the students to do the experiment in small groups. Although the students are not doing the experiment themselves, the teacher's explanations and questioning lead them to consider the "critical details" of the situation (Viennot, 2003); students may demonstrate their engagement by replying to some questions in unison. The picture at the *top right* illustrates a social practice that the teacher has cultivated, in which students first explore ideas in their small groups, and then come to the front of the room to share and explain their ideas to the class. In the picture, the student is preparing her board work before explaining it to the class. Chinese learners tend to consider things privately before sharing their own thoughts publicly and avoid the possibility of "losing face" (Li, 2009), so this kind of practice is difficult to achieve. The teacher developed the practice over approximately one month in quite a structured way. First, students developed their ideas in the privacy of their groups, and two students were called to the front to present their group's ideas. Over time it became more spontaneous, and the presentations led to dialogue, in which other students questioned the presenters or helped them improve their explanations. At the *lower left*, the teacher reinforces the main points in a short and focused presentation. During a discussion driven by students' ideas there often are blind alleys and multiple explanations; although the students mostly figure out the correct explanations, there is a danger that minor points became salient for students or that flawed explanations are remembered. Thus, the teacher reviews what has been accomplished and what is most important to remember; in this, she is using the ideas and writing contributed by students as the raw material for her own explanations. Finally, the *lower right* shows something typical of Western classrooms. There is no obvious focal point in the classroom, and many things are going on simultaneously. At such moments we typically find the teacher assisting students (e.g., getting materials for them, help to fix problems).

With this brief depiction I want to suggest that a learning environment developed in which student ideas and working to getter to develop and improve explanations were central. In this environment, students had much work to accomplish independently, but the teacher played an essential role in providing the supports needed to make their efforts worthwhile (Hmelo-Silver, Duncan, & Chinn, 2007). These developments occurred in both classes and are important to knowledge building. However, there were two problems. (1) In Class A it was difficult to connect work on Knowledge Forum with the classroom discourse; students worked on Knowledge Forum after school and the database was not accessible to them during their small-group discussions. This problem was addressed for the second cohort by installing an IWB, and making one laptop computer available during class to each small group for reviewing the Knowledge Forum database. (2) The approach required considerably more time, and made it more difficult to cover the curriculum. Perhaps fuller integration of Knowledge Forum and classroom discourse may address this problem to some extent, if some of the work to develop explanations can be moderated and completed successfully in Knowledge Forum.

## Descriptive Statistics

Descriptive statistics are shown in Table 1.

Table 1: Descriptive Statistics

| Measure | Class A (2007 to 2008) | | Class B (2008 to 2009) | |
|---|---|---|---|---|
| | M | SD | M | SD |
| FCI pretest | 32.8 | 18.7 | 29.6 | 12.8 |
| FCI posttest | 47.0 | 15.5 | 47.1 | 21.0 |
| FCI normalized gain | 0.16 | 0.33 | 0.26 | 0.28 |
| Content Knowledge | 55.6 | 18.1 | 42.7 | 16.3 |
| Everyday Applications | 61.5 | 14.6 | 59.2 | 19.8 |
| Self Awareness | 66.1 | 18.1 | 61.6 | 17.0 |
| Graphing Task (bicycle) | 53.6 | 19.2 | 36.7 | 11.3 |
| Explanation (N3) | 50.0 | 31.6 | -- | -- |
| Innate Ability posttest | 3.28 | 1.25 | 3.77 | 1.26 |
| Simple Knowledge posttest | 4.23 | 0.91 | 4.17 | 1.05 |
| Quick Learning posttest | 2.59 | 0.93 | 2.81 | 1.12 |
| Attitude to Science posttest | 4.17 | 0.45 | 3.69 | 0.57 |
| Notes Created | 21.5 | 9.33 | 10.2 | 4.88 |
| % Notes Linked | 70.9 | 11.7 | 68.0 | 18.3 |
| % Notes Read | 5.81 | 2.94 | 18.3 | 16.6 |
| Views worked in | 9.14 | 2.29 | 4.73 | 0.90 |
| Revisions | 4.79 | 4.64 | 5.63 | 8.21 |
| Scaffold Uses | 20.3 | 27.1 | 8.78 | 4.95 |

The following observations can be made: (1) The FCI results were consistent with previous research for traditional teaching. For example, Hestenes et al. (1992) studied results from 612 Arizona high school students and obtained a mean pre-test score of 27% and a posttest score of 48%. In a meta-analysis including 1113 high school students, Hake (1998) referred to normalized gains (gain divided by maximum possible gain) as small. However, Hestenes et al. also found that posttest results frequently exceeded 60% if an "active learning" approach was used (normalized gain > 0.45). Suppapittayaporn, Emarat, and Arayathanitkul (2010) also obtained a normalized gain of 0.45 for Grade 10 students in Thailand who used an approach that combined peer learning with structured inquiry. These studies suggest that although the normalized gains that we observed were consistent with traditional teaching, larger values should be possible with a research-based active learning approach like knowledge building. (2) The results for the three learning reflection scales and two in-class tasks are included mainly to pilot test these instruments. They had appropriate properties for use in future design experiments (e.g., mean, SD, and discrimination). In both classes, mean scores were highest for Self-Awareness and lowest for Content Knowledge. (3) In both classes, epistemic beliefs ranged from Innate Ability (strongest) to Quick Learning (weakest). This finding is somewhat surprising because the literature suggests that Chinese learners do not have strong beliefs in native ability, but rather believe in effort Li, 2009). (4) Indexes of participation in Knowledge Forum in Class A were low compared to other studies (Niu & van Aalst, 2009; van Aalst, 2009), which suggests that the use of Knowledge Forum was not optimal. The improvement in the Percentage of Notes Read from 5.81% to 18.3% can be attributed to the new strategy to use laptop computers during class to access Knowledge Forum. Other differences between the cohorts are spurious due to the later start of Knowledge Forum.

## Analysis of Correlations

To test for predictive relationships among the variables, the measures of contribution to Knowledge Forum were combined using exploratory (principal components) factor analysis, and path models were constructed for each cohort. For Class A, Notes Created, Percentage of Notes Read, Views Worked In, Revisions, and Scaffold Uses were factorable into a single factor, *ATK Productivity* (KMO = .771), which explained 62.8% of the variance of the five included variables. For Class B the indicators were less factorable (KMO = .618). There were two factors that explained 69.5% of the variance, the first being similar to ATK Productivity, and the second having the highest factor loading for Percentage Notes Linked. But the second factor was not clearly interpretable as some of its marker items had factor loadings exceeding 0.30 on the first factor. This result suggests that contributions to Knowledge Forum in Class B were more complex than in Class A—although less extensive. Only the ATK Productivity factor was used for the path models.

Due to the small sample sizes the path models are used only in an *exploratory* sense, and only the most promising predictive relationships are reported. These need to be considered suggestive and would need to be verified in a study of a substantially larger sample. The most interpretable results were as follows:

- *ATK Productivity* was the strongest predictor of *Content Knowledge* in Class A (weight $w = 0.54$, $p < .001$), but it did not predict the *FCI normalized gain* ($p > .10$). This suggests that writing notes with ideas relevant to the subject made a strong contribution to general content knowledge, but was insufficient for gains in detailed conceptual knowledge, such as measured by the FCI. FCI gains were only a weak predictor of content knowledge ($p < .10$), which suggests that the two measures captured different aspects of knowledge. In Class B, ATK Productivity did not predict Content Knowledge or FCI normalized gains, but this can be attributed to the lower ATK Productivity scores.
- In Class A *Content Knowledge* ($w$=-0.48, $p < .001$), and in Class B *Everyday Applications* ($w = -0.43$, $p < .01$) were strong inhibitors of *Self-Awareness*. This can be explained if high-performing students on Content Knowledge/Everyday Applications were less reflective—perhaps more sure of themselves—than low-performing students. Verbal responses from the Graphing Task support this interpretation.
- Only weak relationships between the epistemic belief measures and other variables were found. In Class A, *Quick Learning* predicted a decrease in *Attitudes towards Science* ($w = -0.35$, $p < .05$), and *Innate Ability* predicted performance on the N3 explanation task ($w = 0.33$, $p < .05$). The first result is expected because physics does not involve quick learning; the second result may again indicate that students who were confident in their knowledge and attributed this to their ability were more likely to be able to provide correct explanations. In Class B, epistemic belief measures did not predict any other measure, but the poorer result on the Graphing Task was a strong predictor of slightly less favorable Attitudes toward Science ($w = 0.49$, $p < .001$). Thus, limited success on academic tasks early in the unit may be a stronger inhibitor of attitudes towards science than epistemic beliefs.

## Digging into the In-Class Tasks

The idea with the in-class tasks is that they can easily be incorporated into teaching and provide additional data points regarding students' knowledge, explanations, and problem solving as a unit unfolds. It is therefore worth looking more deeply into what they reveal about student thinking. The following informal analyses were completed only for Class A.

For the *Graphing Task*, scores from the graphs were used to divide the class into below median and above median groups. Verbal responses were then analyzed for each group, using the following categories: procedural, graph feature, conceptual question, higher-order thinking, and problem posing. For example, 'procedural' referred to procedures such as how to use motion detectors and formulas, and 'graph feature' indicated that a student had correctly described a graph feature. A question was coded higher order thinking if it sought understanding at a conceptually higher level, for example, if a student said s/he needed a clearer understanding of the relationships among the graphs or introduced a conceptual issue that had not been considered before, such as whether a vertical section of a *v-t* graph could be exactly vertical. Responses for "I would like to raise …" were coded *problem posing* if they raised a conceptual problem.

The results indicated that: (1) Above-median students wrote shorter responses (40 words on average, compared with 60), and more often used formulas or symbols. (2) Almost all students provided at least one correct description of a graph feature. (3) The below median asked many conceptual questions (15, compared with 7 for the above median group). This result is encouraging because it suggests that below-median students were engaged with the concepts to be learned, even if they did not understand them deeply at the time. (3) Above median students made more procedural statements and problem statements. (4) Both groups asked few statements coded as higher-order thinking.

Students' responses to the *N3 Explanation Task* were divided into four categories: agreement with the horse (incorrect); disagreement without providing an analysis based on physics principles, or with an incorrect analysis; disagree with a mostly correct analysis (minor errors), and disagree with a fully correct analysis. Results showed that almost all students came to the correct conclusion (which is common sense), but relatively few used physics principles to construct their argument (14 of 41 students). Nevertheless, ten students provided a completely correct argument. Although these results are somewhat disappointing for an active learning approach—the issue that was probed is fundamental to understanding Newton's third law—there was scope for the diffusion of accurate knowledge within the class. Knowledge Forum can facilitate this diffusion.

## Conclusion

The goal of this study was to explore how knowledge building could be developed in a way that is sensitive to Asian values and practices. It was argued briefly that although the influence of Confucianism has been declining in the 20th century, the major goals of education in Asia continue to be self-improvement and upward social mobility, and that the HKCEE is a difficult examination. To cope with the level of difficulty, students use an approach that combines efforts to understand with memorization and extensive practice, and teachers reserve extensive instructional time for practice for the examination (Chan & Rao, 2009; Gao, 1998). In this kind of cultural context, although students may interact with each other, it is unlikely that the teacher would implement a primarily activity-based approach—whole-class teaching and teacher guidance remain important (Wu & Huang, 2007). The four pictures in Figure 1 depicted the mix of teacher guidance and student agency achieved in the two classes of this study. The approach the teacher developed included didactic whole-class instruction, eliciting the ideas of students, student-student talk, the teacher's work to emphasize critical details, and independent work by students. It is proposed that this environment is a significant step towards an interactive learning environment suitable for Hong Kong schools.

The study used a variety of measures to study students' performance, and led to many insights into how knowledge building may be developed or studied in Asian contexts. I briefly reiterate three points.

First, although result in terns of conceptual knowledge were acceptable in the setting studied, and consistent with research on traditional physics, future work must aim for larger gains. There are good reasons to think that mean posttest scores of approximately 60% are possible (Suppapittayaporn, et al., 2010; Hake, 1998; Wells et al., 1992). As a PER-based approach, knowledge building must ehnace the extent to which students develop conceptual knowledge. The instructional apprioach was focused on conceptual knowledge, but the processes started in the classroom may require extension. It is hypothesized that further development of the approach the teacher began to implement in Class B, which involved creating a more seamless learning environment in which the use of Knowledge Foeum was integarted with classroom events, is nececsary.

Second, the study suggested how this may be accomplished. Productivity in Knowledge Forum (writing and reading notes) made a significant contribution to general physics knowledge, but not to conceptual knowledge. As the path weight for productivity in Knowledge Forum was relatively high, it may not be possible to enhance conceptual knowledge by increasing productivity alone. It is hypothesized that a direct effect from Knowledge Forum on conceptual knowledge may be possible if relational aspects of work in Knowledge Forum are increased (i.e., notes that establish linkages, synthesize progress, and identify rise-above ideas). This would require explicit attention to the development of sophisticated uses of Knowledge Forum (van Aalst, 2006).

Finally, the study showed that these Asian students did have epistemic beliefs, but that these did not interact with the variables studied, particularly contributions to Knowledge Forum. One possible explanation is that other beliefs such as filial piety masked such interactions. But it is also possible that the level of work on Knowledge Forum and other activities were not sufficient to cause changes in epistemic beliefs. Further research would be useful to shed additional light on this question.

## References

Aldridge, J. M., & Fraser, B. J. (2000). A Cross-Cultural Study of Classroom Learning Environments in Australia and Taiwan. *Learning Environments Research*, *3*, 101-131.

Amano, I. (1990). *Education and examination in modern Japan*. Tokyo, Japan: University of Tokyo Press.

Biggs, J. B. (1996). Western misperceptions of Confucian-heritage learning culture. In D. A. Watkins & J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 45-67). Hong Kong: CERC and ACER.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations* *in leaning: New environments for education* (pp. 289-325). Mahwah, NJ: Lawrence Erlbaum Associates.

Chan, C. K. K., & Rao, N. (2009). The paradoxes revisited: The Chinese learning and changing educational contexts. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: Changing contexts, changing education* (pp. 315-349). Hong Kong: Springer/Comparative Education Research Centre.

Education and Manpower Bureau (EMB) (2000). *Learning for life, learning through life: Reform proposals for the education system in Hong Kong* (Report).

Gao, L. (1998). Cultural context of school science teaching and learning in the People's Republic of China. *Science Education, 82*, 1-13.

Gao, L., & Watkins, D. A. (2002). Conceptions of teaching held by school science teachers in P.R. China: Identification and cross-cultural comparisons. *International Journal of Science Education*, *24*, 61-79.

Hake, R. R. (1998). Interactive engagement versus traditional methods: a six-thousand-student survey of meachanics test data for introductory physics courses. *American Journal of Physics, 66*, 64-74.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141-158.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*, 99-107.

Lee, W. O. (1996). The cultural cntext for Chinese learners: Conceptions of learning in the Confusian tradition. In D. A. Watkins & J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 25-41). Hong Kong: CERC and ACER.

Leung, F. K. S., Yung, B. H. W., & Tso, A. S. F. (2005). Digging below the surface: Secondary analysis of TIMSS data for Hong kong and its implications for educational reform. *Canadian Journal of Science, Mathematics, and Technology Education, 5*, 329-360.

Lewis, R., Romi, S., Qui, X., & Katz, Y. J. (2005). Teachers' classroom discipline and student misbehavior in Australia, China and Israel. *Teaching and Teacher Education, 21*, 729-741.

Li, J. (2009). Learning to self-perfect: Chinese beliefs about learning. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: Changing contexts, changing education* (pp. 35-69). Hong Kong: Springer/Comparative Education Research Centre.

Niu, H., & van Aalst, J. (2009). Participation in knowledge-building discourse: An analysis of online discussions in mainstream and honours social studies courses. *Canadian Journal of Learning and Technology/La Revue Canadienne de l'Apprentissage et de la Technologie [online], 35*(1). Retrieved from http://www.cjlt.ca/index.php/cjlt/article.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago, IL: Open Court.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97-115). New York, NY: Cambridge University Press.

Schraw, G., Bendixen, L. D., & Dunkle, M. E. (2002). Development and validation of the Epistemic Belief Inventory (EBI). In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 261-275). Mahwah, NJ: Lawrence Erlbaum Associates.

Suppapittayaporn, D., Emarat, N., & Arayathanitkul, K. (2010). The effectiveness of peer instruction and structured inquiry on conceptual understanding of force and motion: a case study from Thailand. *Research in Science & Technological Education, 28*(1), 63-79.

van Aalst, J. (2006). Rethinking the nature of online work in asynchronous learning networks. *British Journal of Educational Technology, 37*, 279-288.

van Aalst, J. (2009). Distinguishing knowledge sharing, construction, and creation discourses. *International Journal of Computer-Supported Collaborative Learning, 4*, 259-288.

Viennot, L. (2003). *Teaching physics*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Wallace, J., & Chou, C.-Y. (2001). Similarity and difference: Student cooperation in Taiwanese and Australian science classrooms. *Science Education, 85*, 694-711.

Wu, H.-K., & Huang, Y.-L. (2007). Ninth-Grade Student Engagement in Teacher-Centered and Student-Centered Technology-Enhanced Learning Environments. *Science Education*, *91*, 727-749.

## Acknowledgments

This research was supported by two grants from the Research Grants Council of Hong Kong (Grant Nos. HKU 752508H and HKU 741007H). The author thanks the teacher and students for their hard work and opening their classroom to this classroom research, and to Katerine Bielaczyc, Carol Chan, Jianwei Zhang, and John Gilbert for their visits to the classroom and their comments on what they observed. Permission to publish the photographs in Figure 1 is also gratefully acknowledged.

# Using collaborative activity as a means to explore student performance and understanding

Marcela Borge, John M. Carroll
Pennsylvania State University, College of Information Sciences and Technology, University Park, PA 16802-6823
Email: mborge@psu.edu, jmcarroll@ist.psu.edu

**Abstract:** This paper presents data from the beginning process of restructuring a usability-engineering course to incorporate aspects of collaborative-process theory with required course content. Students' collaborative processes were evaluated at the end of the semester in order to identify key areas that instructors would need to support better in future iterations of the course. Overall students were skilled communicators, but lacked ability in areas of planning, productivity and critical evaluation. A closer examination was made of difficulties that students had with critical evaluation and trade-off analysis processes since these were key aspects of the course. Findings from analyses were used to suggest ways that the course could be modified in the future to better prepare students to address these important concepts in usability engineering.

## Introduction

As more emphasis in technology and science classes gets placed on getting students to do well on product-based outcomes, less emphasis is placed on getting students to understand and engage in important learning processes. This is extremely problematic as researchers and field practitioners in science education have emphasized the growing need for and increasing lack of students' abilities to understand and engage in critical processes associated with collaborative intellectual-activity: the ability to communicate effectively, listen and build on ideas, negotiate ideas, and engage in scientific argumentation (Barron, 2003; Driver, Newton, & Osborne, 2000; Felder, Woods, Stice, & Rugarcia, 2000). Such processes, though identified as critical in educational and work contexts, have been identified as a weakness among American students (SCANS, 1991). The attention focused on collaborative skills has led many schools to increase the number of team-based projects, and collaborative activities they require students to participate in. A fundamental problem with this response is that it assumes that students already possess the cognitive and social skills necessary to engage in effective collaborative endeavors and that these competencies will naturally emerge and develop through exposure to collaborative learning environments; this is wrong. Research has shown that individuals are often frustrated by group interactions, waste time (Salomon & Globerson, 1989), or fall victim to a variety of other problems that lead to dysfunctional group-processes (Webb & Palincsar, 1996). This suggests that students need more explicit guidance with collaborative endeavors than is currently typical.

Gaps in knowledge about what constitutes effective collaborative interactions and/or the ability to apply these concepts in practice are particularly problematic at the college level where individual grades may depend on a team's performance. In The College of Information Sciences and Technology (IST), where our participants have spent the last four years, students are expected to work in teams for much of their academic careers, depending on each other for grades based on their collaborative efforts. Given that dysfunctional interactions among team members can lead to many negative outcomes such as lower performance, alienation, failure, and even the likelihood that students will drop out of the group, the course, even the major (Barron, 2003; Hogan, 1999; Light 1990; Rosser, 1998), it stands to reason that over a four-year time span such trends could lead to considerable problems for these students, and many faculty regularly see these problems play out day-to-day. In order to diminish these problems it is necessary to develop students' collaborative competence. Toward this aim, the authors are currently undertaking a restructuring process for a usability-engineering course in the College of Information Sciences and Technology (IST). The goal for the course is to help students understand and apply important concepts of usability engineering and collaborative-competence theory simultaneously. However, time constraints, the breadth of domain knowledge, and student push back make this an uphill battle. This paper represents the beginning of our journey down Dillenbourg's long road (Dillenbourg & Traum, 1999): moving beyond distribution of the task to a shared understanding of it. The work presented here is a specific report based on one small strand of our research that encompasses a much broader data set.

### The usability-engineering course

Usability engineering is defined as the concepts, processes, and practices engaged in the process of developing software systems and applications to ensure that they serve their intended users effectively. To a considerable extent, it mirrors and complements software engineering. Our approach to teaching usability engineering relies on user interaction scenarios as the primary context for analyzing requirements, describing specifications, envisioning designs, developing rationales, creating various sorts of prototypes, and evaluating systems and applications. Our usability-engineering course is organized around a semester-long system development project. The course works its way through a curriculum organized by the flows of the system development process, and student teams define, implement, and test their projects respectively throughout the course. This course has been previously successful with computer and information science students. In 2002, a textbook based on the course was published (Rosson & Carroll, 2002; see http://ist413.ist.psu.edu/ for most recent syllabus and class-by-class activities).

When we designed the usability-engineering course, we carefully devised a list of six main learning goals for our students and came up with various ways that we could implement and assess them (Carroll & Borge, 2007). In this paper we will evaluate students with regards to one of these goals: the ability to build knowledge base through collaborative discussions and refine necessary skills in collaborative groups.

## Four core collaborative capacities

Four core capacities within the realm of collaborative learning were identified from educational research literature (Borge, 2007) and unpacked for students. These capacities were intended to embody an organized theory of goals and objectives students would need to meet in order to work in an effective collaborative environment. These capacities were introduced and incorporated into the students' collaborative work. These capacities were originally created as an attempt to structure group processes to address the many factors that can contribute to group-process problems (Borge, 2007; White & Frederiksen, 2000). Originally presented as roles, they were also a useful way to simplify a complex endeavor, such as collaborative interaction, into more manageable sub-skills that students could learn about, demonstrate to each other, and improve upon through practice (Brown & Palincsar, 1985).

The four collaborative capacities are planning manager, communication manager, critical evaluation & negotiation manager, and productivity manager. Physical tools and activities were developed to support the learning and use of the roles/capacities. One of the primary tools was a guide for students that detailed each capacity: the major objectives, problems students may need to prevent/correct, and strategies that they could use to prevent/correct problems (for examples of the guides see http://ist413.ist.psu.edu/, week-to-week activities, calendar, 2/4, under collaborative roles). These Guides were originally developed for children ages 10-14 years of age, but extensively modified for use in this course for use with usability contexts and college age learners. We used feedback from instructors and students to tailor the language, problems, and strategies to specifically address problems faced by this population of students (Borge & White, 2009). Examples of common problems for this population include: improper use of time during team meetings, lack of critical evaluation, and lack of accessibility to collaborative products in real-time. The updated guides included information to prevent and/or correct such problems.

## Research methods, objectives, and class context

The research objectives for this project were to fundamentally improve how team activities were implemented, supported and structured in a usability-engineering course so as to improve students' ability to both understand and apply important concepts. Since application of concepts is a critical aspect of the research we will focus on the analyses of an in-class activity requiring the application of usability engineering concepts through particular collaborative interactions. We used assessments of students' collaborative processes from these videos and compared them to their final project grades to see whether students' process performance was correlated to students' performance on their team project.

Participants were students enrolled in a senior level usability-engineering course (IST413). Students were divided into eight teams of five-to-six students: two teams composed of two females and four males and six composed of all males. The teams worked together throughout the semester on collaborative in-class activities, collaborative homeworks, and a semester-long usability-engineering project. All students were required to learn about effective collaborative-interactions, and practice applying goals and strategies during these activities. Students learned about the four capacities, their objectives, and ways of promoting effective collaborative interactions (i.e., setting process goals, and using strategies to prevent or correct team-problems). Goals and activities related to the four capacities were incorporated with 13 in-class activities during which students were expected to work together and turn in a deliverable (an activity worksheet). The activities were either tied to the content covered that week, or to the students semester long project. The roles were a required part of the course since activities, collaborative homework assignments, and even some quizzes required students to practice and understand the roles, their goals and strategies. Students' overall collaborative capacities were assessed from video of teams engaged in an in-class activity in week 13 of the course (for details of the activity see http://ist413.ist.psu.edu/, week-to-week, calendar, activity 13 on 2/28). This videotaped activity was a brainstorming and decision-making session that required students to use and apply concepts they had learned throughout the semester: write design scenarios, identify potential user problems, suggest ways to resolve problems with documentation/help design, evaluate trade-offs of proposed designs, etc. We assessed students on their collaborative interactions by using a scoring rubric developed by Borge, 2009 (See Table 1 for the four capacities and objectives that students were assessed on). This rubric is directly connected with the collaborative-capacity guides as the objectives they assess are those proposed in the guides. Out of eight teams seven consented to be included in our research; these teams were used in analyses. There were five objectives for each of four capacities, for a total of 40 possible points. However, one objective in the planning capacity, and one objective in the productivity capacity were not applicable to this activity (those appear shaded in Table 1). Therefore, there was a total of 38 possible points on this measure.

Table 1: The four capacities and their objectives. Objectives not assessed appear shaded.

| Capacities | Objectives |
|---|---|
| 1. Communication | - Most members participate during discussions and quiet members are encouraged to contribute<br>- Team members listen intently<br>- Team members build on each other's ideas<br>- The team works to create a common understanding<br>- Comments are clear, concise, and relevant to the task at hand |
| 2. Planning | - There is a set agenda for the meetings<br>- The team has a long-term plan for the entire project<br>- The team has a set way of providing each other with feedback<br>- The team has set project/interaction goals<br>- The team assigns tasks and responsibilities to ensure that everyone makes equal contributions to the project |
| 3. Critical Evaluation/ Negotiation | - The team considers multiple points of view from various members<br>- The team critiques ideas respectfully and knows difference between positive and negative conflict<br>- The team takes time to evaluate trade-offs of different ideas<br>- The team takes all members into account when deciding which ideas/courses of action to take.<br>- The team keeps track of the ideas that are presented, discussed, and evaluated |
| 4. Productivity | - There is a set method for determining and recording progress<br>- The team does not display off-task behaviors<br>- The team evaluates and works to improve work quality<br>- The team sets time to update each other of their progress<br>- The team turns in high quality deliverables in a timely manner |

Students were familiar with the rubrics, as they had used them to assess themselves at the beggining of the semester. When students completed a video-taped activity, at the begining of the semester, they were asked to watch their videos, score their team with these rubrics, and identify areas that seemed particularly problematic for their team. The scoring system was as follows: 0 = does not fuilfill objective to any degree, 1= has some ability to fulfill objective, but still somewhat problematic, 2= fulfills objective perfectly, could be used as ideal example for other students. Once they identified problem areas they were asked to look through the guides and select strategies to prevent/correct the problems they identified. This activity was used as a means to introduce the roles, their guides, and ways to apply the concepts they presented. It was also a means to familiarize students with the rubric we would use to assess their interactions at the end of the semester. At the end of the semester, students were told that we would assess them with the same rubrics they used at the beginning of the semester and we repeated the same procedures as during the first recorded activity.

## Reliability of the scoring rubric

The results we report are derived from overall scores determined by the first author. however, we assessed the overall reliability of the scoring rubric by training a graduate student to be a second rater. The second rater was presented with the same materials and guides the students were given. Twenty percent of the data set was doubly coded with the following results: Kappa=.69, and r=0.88, n=140, p< .001. Transcripts were not available at this point in the analysis, so the coding was based directly on video clips. The two coders matched exactly on 78.8 % of the items. Most of disagreement regarded teams for which the first author identified modest critical evaluation (score of 1), but the graduate-student coder felt there was no critical evaluation demonstrated (score -9, n/a).

## Transcription of video

Each of the seven 30-minute videos was transcribed following a similar format. Each new speaker utterance and or behavior was numbered, denoting a new "turn". A "turn" ended when a different speaker introduced a new utterance. The students were given pseudo-names in the transcript. Parentheses were used to label nonverbal gestures and events (i.e., leaving the group, making faces, using hands, etc). Brackets were used for codes, time stamps, and notes relevant to the analyses, but not found in the video itself. These transcripts were utilized when the author scored each team's collaborative interactions with the given rubric, and for microanalysis of processes.

## Results

### Teams' collaborative interactions: overall findings

Out of 38 possible points in the assessment, the average score across all seven teams was 50% and the median score was 47.3%. This means that all teams scored in the "average" range of performance, getting mainly scores of "1" on the different objectives in the assessment. Scores of "2" were far less common, except in the communication category. Collaborative performance was not significantly correlated to measures of individual performance (i.e., exam scores, class participation, or overall course grade). However, the collaborative assessment was correlated to performance on measures assessed at the team level: where students worked together on one product and shared the same grade ($r$=.77, $p$<.05). This suggests that improving collaborative performance could lead to improved performance on deliverables created and assessed at the team level.

When we examined how teams performed on each individual capacity, we found that they performed significantly better on one of them. There was a significant difference between how students performed on the communication capacity ($M$=.70, $SD$=.18) and how they performed on the assessment as a whole ($M$=.50, $SD$=.13), $t(6)$=6.8, p<.001). Communication was the students' strongest area reflecting strengths in listening for understanding and building on each other's ideas. In fact we saw very little evidence of parallel talk, where one utterance, or turn, is followed by a different and unrelated utterance.

Table 2. Descriptive statistics of teams' overall performance on capacities:
crit = critical evaluation, com = communication, pro = productivity, and plan = planning.

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| crit | 7 | .30 | .80 | .5143 | .19518 |
| com | 7 | .40 | .90 | .7000 | .18257 |
| pro | 7 | .25 | .75 | .4464 | .17466 |
| plan | 7 | .25 | .63 | .4107 | .13909 |
| Valid N (listwise) | 7 | | | | |

There were no significant differences between the rest of the capacities and overall performance with students scoring in the average range for all three capacities. Particular aspects of the planning capacity were not evident in student behavior. For example, students largely accomplished the activity problem by problem, with no thought as to what previous information might be helpful, what the goals of the activity were, or how they would ensure that they achieved those goals. Some aspects of productivity were also missing from students' interactions. Even though they were quite good at staying on task and progressing through items quickly, they still did not perform above average in this capacity. Most of their low scores were due to the fact that they 1) did not show evidence that they had set methods for determining and recording progess and 2) failed to evaluate work quality. Many aspects of this activity depended on students pulling on previous assignments and, for the most part, team members were unaware of who was responsible for which tasks. Failure to evaluate work quality was also a problematic and reaccuring theme. Most teams only cared that the answer was "good enough", not that it was the best possible answer. In fact, student teams actually used the term, "good enough" an average of two times during the activity.

The last capacity we would like to discuss is the critical evaluation/ negotiation capacity. Given that many parts of the activity explicitly asked students to engage in evaluation processes, we would have expected to see higher than average scores in this area, but were disappointed to find that only three teams were able to score any 2s in this area, and only one team fulfilled the trade-off analysis objective, demonstrating the ability to effectively apply the concept of trade-off analysis (a process we will discuss later). All other teams scored 0s on this objective.

### Specific problems with the capacity for critical evaluation/ negotiation

Thre was no need for conflict managment or negotiation of differing points of view as students did not seem particularly invested in critiquing or evaluating ideas. Students were very respectful to fellow teammates and there were no instances of intimidation or personal attacks during collaborative activities. Unfortunately, students seemed to be crossing off mental checklists rather than engaging in deeper forms of analyses or argumentation. In fact, students accepted or built upon ideas without stopping to evaluate or challenge those ideas or ask team members to provide rationale for suggestions. Questions, such as, "why do you think we should that" simply were not evident in the team interactions. For the most part, students seemed to lack the ability to step back from problems and evaluate their suggestions as objects of thought. One of the best examples of students' inability to take a step back and think about a problem and its solutions critically came from one of the teams presented below. This particular team was creating a website for a photo club. In this example, the team is working on the second question in the activity.

> *Question 2. Discuss the design scenarios you have prototyped. Choose 2 that you think are most likely to raise problems for learning or use, i.e. where help or other documentation might be required for at least some users. Summarize each scenario briefly, and indicate why you think it might be a candidate for help or training.*

The team (M1, M2, M3, M4, and M5) tries to come up with problems that a user could experience when trying to use their website. Since their website is for a photo club, M3, suggests that they might have problems uploading a picture file. M4 asks M1 to log into the website in order to check for restrictions. M1 tries to log on, but does not remember the password. The rest of the team continues to talk about the potential problem (see Table 3).

Table 3: Team talk as they work to find a solution to a problem in the activity.

1. 00: 20: 24.22 M3: How many picture files can you use?
2. 00: 20: 25.14 M2: It was on the website...
3. 00: 20: 25.25 M1: Yeah where's that at...
4. 00: 20: 33.10 M2: Yeah go to the homepage... yeah, there you go...
5. 00: 20: 35.28 M1: What's the password, anybody...
6. 00: 20: 40.28 M3: Um the other Mike has that, but he's not here... I don't know the password of---
7. 00: 20: 47.13 M4: Why couldn't he make it something easy like one two three four...
8. 00: 20: 49.15 M3: He made it something easy but it was just like---
9. 00: 20: 51.14 M1: Penn State?
10. 00: 20: 52.19 M2: Try it.
11. 00: 20: 54.29 M4: Did he actually send you the---
12. 00: 20: 56.12 M1: No, I just remembered it was something stupid.
13. 00: 20: 57.22 M2: He said it out loud to us.
14. 00: 20: 58.02 M4: Oh... I can't remember!
15. 00: 21: 01.19 M2: Security question?
16. 00: 21: 06.23 M3: Um?... Calendar help training solution, we could just provide like a document that showed you how to do stuff.
17. 00: 21: 26.06 M1: Yeah I'm not sure this is the right, its asking for yahoo ID.
18. 00: 21: 32.05 M3: Even if just [inaudible] the document that shows them where the help is for Flickr.
19. 00: 21: 36.17 M2: That's gonna have a [Inaudible].
20. 00: 21: 37.25 M4: Flickr for beginners?
21. 00: 21: 41.11 M2: Or what happens if you go to the picture page and click on a picture?
22. 00: 21: 50.22 M1: It's asking to, ah, open up.
23. 00: 21: 58.24 M2: [Inaudible] pictures so...
24. 00: 22: 03.05 M1: Um.
25. 00: 22: 07.29 M2: Oh wait.
26. 00: 22: 11.21 M3: What is it?!

The team continues discussing possible problems with uploading pictures in turns 8-11. Meanwhile, M1 still cannot remember the password and asks the team for help (turn 12). In the next 11 turns the team tries to guess the password to no avail. Nonetheless, in turns 25-26, the team marches on, trying to come up with another idea for a problem that a user could experience when trying to use their website, while M1 continues to try to log on. The team proceeded in this manner for two minutes. Unable to resolve the problem and log on, one of the students suggested they simply summarize their previous scenario about a user experiencing problems uploading pictures, but this time add an FAQ or search page to help the user resolve the problem.

What happened here? It seems these students were so engaged in trying to verify a proposed solution to the problem, that they completely ignored the most obvious one: the problem they were experiencing at the moment of not being able to login because no one could remember the password. Forgetting the password is a common problem that many sites simply resolve by allowing you to reset the password. However, for this particular website that solution would not work because all the users share the same password. This was a trade-off that came up in their design, they wanted all the users to be able to link Google Calendars and other aspects of the website to Facebook, but the only way that could work was for all of them to share the same password. This solution gave them want they wanted in a feature without them having to write programming code. However, it had drawbacks as their transcript demonstrated. Identifying these types of trade-offs and thinking ahead for possible solutions was the main goal of the activity, but sadly, not one this team could accomplish. This team was not alone.

## Trade-off analysis

Trade-off analysis is a fairly complicated critical-evaluation process that is also a core concept of usability engineering. It involves weighing design goals (i.e., the needs of your users, ease of use, added features, etc.) with constraints and resources of the design team (i.e., cost to build and maintain, programming demands, client demands, etc.) in order to make the most reasonable decisions. Trade-offs, by definition imply that there is no correct answer, just the best one given design goals and resources; it is a balancing of factors that cannot all be attained simultaneously. Given its centrality to this course it was one of the critical evaluation objectives that students were assessed for as part of the collaborative assessment. This was the only objective in the assessment in which teams received zeroes across the board. Only one team, Team 6, was able to effectively apply the concepts inherent to trade-off analysis. This team will be used to illustrate the difference between how the majority of students engaged in the trade-off analysis process, with what was expected.

In question three of the activity, the students were asked to "Analyze trade-offs associated with your design idea. Consider a variety of issues, e.g. cost to build or maintain, generalization to other scenarios, reactions by users, and so on". Six out of seven teams simply responded by coming up with a list of pros and cons. The teams did not justify their design or why it was the best possible solution. They also did not weigh factors involved in their decision making. Instead, team discussions were more akin to identifying possible problems. The following example is representative of the pattern of interactions that stemmed from teams trying to engage in the process of trade-off analysis. In this team, members (M1, M2, M3, M4, and M5) have just completed question two (presented with the example in Table 3) and just now reading question three (see Table 4).

Table 4. Team talk representative of the majority of teams as they worked to resolve Question 3.

| | | | |
|---|---|---|---|
| 1. | 00:18:34.05 | M1: | can you read the question |
| 2. | 00:18:34.07 | M2: | It just says analyze the trade offs, consider a variety of situations |
| 3. | 00:18:43.29 | M1: | [inaudible] ... you can say there is almost no way... if it happens it's going to come down to the officers of DDF [Dance Dance Fanatics] to do something |
| 4. | 00:18:59.02 | M3: | Yeah |
| 5. | 00:18:59.03 | M4: | Yeah |
| 6. | 00:19:03.00 | M3: | Like when you sign-up you should be required to like... accept--- |
| 7. | 00:19:08.13 | M1: | Like one of those... agreements [inaudible] |
| 8. | 00:19:10.29 | M3: | yeah, that you always accept and just push ok. |
| 9. | 00:19:17.18 | M4: | Something with the calendar page |
| 10. | 00:19:22.09 | M1: | Um |
| 11. | 00:19:45.16 | M2: | Members using calendar may f[**] it up. |
| 12. | 00:19:52.00 | M1: | Ha, yes! |
| 13. | 00:19:54.21 | M3: | He stole the words outta my mouth! |
| 14. | 00:20:01.21 | M2: | They... what's a better word than that?! |
| 15. | 00:20:07.22 | M1: | May not fill in all the field, or something like that. |
| 16. | 00:20:08.05 | M2: | May hinder--- |
| 17. | 00:20:10.12 | M1: | --- So their, maybe events on there but they may not have like time or place or combination of something or description just be like meeting |
| 18. | 00:20:21.17 | M2: | Like the date and time. |
| 19. | 00:20:22.05 | M1: | Might just be ambiguous as to what the person was talking about. |
| 20. | 00:20:26.10 | M4: | The date is like required. |
| 21. | 00:20:28.12 | M1: | Yeah its like, you'll know the day, but its like it might just be completely ambiguous and just say meeting... [inaudible]. |
| 22. | 00:20:50.28 | M2: | I think we're good with three. |

In turns 3, 9, and 11, Students present two problems with their design: 1) there is no simple way to monitor content, and 2) users may leave important fields in the calendar page blank. In turns four through nine, students unpack the first problem, and in turns 10-21 students unpack and discuss the second. As students speak a member writes down these problems on their team paper. There is no discussion about alternative designs or justification for why they have chosen this design even though these problems are present. In turn 22 they simply state that they have sufficiently addressed question three and move on to question 4. This was the predominant pattern of talk that ensued from 6 teams during this trade-off analysis portion of the activity: identify problem, agree/build on idea, identify another problem, agree/build on idea, write ideas down and move on. Only one team demonstrated any ability to engage in trade-off analysis, Team 6 (see Table 5 below).

Table 5: Team talk of students demonstrating ability to apply trade-off analysis concepts.

| | | | |
|---|---|---|---|
| 1. | 00: 26: 27.22 | M2: | Some others, trade-offs. |
| 2. | 00: 26: 32.11 | M4: | Cost to build or maintain, photo gallery is like it takes time for people to like |

3. 00: 26: 41.18 M3: Update their galleries because they have to send them to the administrator and then the administrator has to go and
4. 00: 26: 44.13 M1: Actually (mumble)
5. 00: 26: 47.28 M3: Yeah
6. 00: 26: 47.20 M4: (Mumble) You got a hundred members and all these photos just gonna be taking up space and all that stuff.
7. 00: 26: 55.07 M1: All, well if we get a hundred members [inaudible---all talking at once].
8. 00: 26: 59.24 M4: I don't know how restrictive the university is with all that though.

Conversation continues...

14. 00: 27: 29.03 M4: That's why it has to be each user can just login and upload their own photos
15. 00: 27: 36.09 M2: Cost to build and maintain.
16. 00: 27: 38.08 M4: So that's one of our trade offs I guess, wouldn't it be? Cause ours is like that, and after--- and not so much functional I guess, because of our own time to build something like that with our limited resources.

The patterns of talk are quite different in this team. They begin by identifying a relevant trade-off, cost to build and maintain (turn 22) and then unpack what the trade-off means as it applies to their project (turns2-8). This type of discussion continues as students build on each other's ideas. Until finally, in turn 14, an alternative design idea is proposed that would address their design problems. At which a member reminds them of the trade-off, cost to build and maintain (turn 15) and another team member summarizes their discussion by stating that this is in fact their trade-off: they are settling for a website with vey little functionality because they lack the time and resources to produce anything more complicated. The majority of teams started by identifying problems with their design and never worked their way up to how they fit into a bigger category of trade-offs or thought about how a trade-off category would play out with their project. Team 6 started with a trade-off, worked their way down by unpacking the trade-off as it applied to their system, identified related problems, alternative designs, and then weighed these designs in order to decide which would work best for their team. They did this by summarizing their discussion and defending their original design idea with concrete rationale.

## Discussion

Findings from this study have helped us to identify problematic concepts for students and provided us with insights as to how to modify the course in future iterations to address these issues. This course contained a group of seemingly engaged students, who appeared to be collaborating well on the surface. However, when their interactions were closely examined we were able to see that students were falling considerably short of what could be attained in terms of collaborative competence. Students in this course were very competent communicators, but were still lacking abilities to critically evaluate their ideas and work quality as well as plan their work and interactions. We also discovered that students' understanding and practice of the trade-off analysis process is not in keeping with the major goals of the course. Our novice usability-engineering students seem to be making mistakes analogous to Schoenfeld's (1989) novice mathematicians; they were picking solutions to problems and not stopping to evaluate if they were the best ones. Thus we should employ teaching methods similar to those used by Schoenfeld to improve students' critical thinking skills as well as their ability to step back from a problem and think about their performance as an object of thought. Our findings also suggest that we can improve team deliverables (the project score) by improving process learning (their collaborative score).

We are currently developing process-learning activities that can better meet the needs of our students for the next iteration of this course. For example, expert thinking will be made more visible for our students (Collins, Brown, & Newman, 1991) in three ways: 1) the instructor will model trade-off analysis for students with the help of colleagues and focus more on explaining his rationale behind usability decisions by weighing options, 2) students will be shown contrasting examples (such as those presented in Tables 4 and 5) of students engaging in trade-off analysis, and 3) students and instructors will work together on collaborative problem solving activities; the aim being to move to a more discussion centered methodology where evaluation of processes are common topics. We are also trying to minimize the "good enough" tendencies in our students by forcing teams to defend their design ideas during presentations to their peers, and also require peer evaluations of project deliverables throughout the semester. Thus motivating them to focus more on their work-quality and their decision-making processes.

### Closing thoughts: the importance of professional collaboration and course evaluation

It is essential for educators to ensure that the collaborative learning opportunities we provide for our students consistently accomplish what we intended them to do: give students opportunities to learn from and challenge the ideas of others, and give them an authentic context to practice applying the core concepts and techniques of a domain. Collaborations between domain experts and experts in the science of learning, such as that between the authors of this paper, are crucial to accomplishing this goal. This type of collaboration is necessary help set appropriate learning goals for students, identify learning goals that present particular problems for students, and implement changes to a curriculum to address these difficulties. Only through these types of collaborations and course evaluations we can fully develop students' learning potential and provide them with richer, more meaningful courses that prepare them to function in an increasingly team-oriented workplace. We do not pretend that these types of domain and process learning fusions do not come with extreme challenges, as we faced many. Among them was a huge push back from students who genuinely felt as though they already knew all there was to know about effective collaborative practice; our finding demonstrate that this is not the case. We contend that the only way to get past students' misconceptions of collaborative practice are to make collaborative goals and processes part of an ongoing conversation between instructors and students that begins in elementary school and continues on throughout their educational and professional careers.

## References

Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences, 12*(3), 307-359.

Borge, M. (2007). Regulating social interactions: developing a functional theory of collaboration. *Dissertation Abstracts International*, 241.

Borge, M., & White, B. Y. (2009). *Scaffolding collaborative processes with managerial roles.* Paper presented at the American Educational Research Association, San Diego, CA.

Brown, A. L., & Palincsar, A. M. (1985). *Reciprocal teaching of comprehension strategies : a natural history of one program for enhancing learning*. Champaign, Ill. Cambridge, Mass.: University of Illinois at Urbana-Champaign; Bolt Beranek and Newman Inc.

Carroll, J., & Borge, M. (2007). Articulating case-based learning outcomes and assessment. *International Journal of Teaching and Case Studies, 1*, 33-49.

Collins, A., Brown, J. S., & Newman, S. E. (1991). Cognitive Apprenticeship: Making Things Visible. *American Educator: The Professional Journal of the American Federation of Teachers, 15*(3), 6-11,38-46.

Dillenbourg, P., & Traum, D. (1999). *The long road from a shared screen to a shared understanding.* Paper presented at the Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference, Palo Alto, California.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*, 287-312.

Felder, R., Woods, D., Stice, J., & Rugarcia, A. (2000). The future of engineering education II. Teaching methods that work. *Chem. Engr. Education, 34*(1), 26-39.

Hogan, K. (1999). Sociocognitive Roles in Science Group Discourse. *International Journal of Science Education, 21*(8), 855-882.

Light, R. (1990). Explorations with students and faculty about teaching, learning, and student life. Cambridge, Mass.: Harvard University Press.

Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: scenario-based development of human-computer interaction*. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Rosser, S. (1998). Group work in science, engineering, and mathematics: Consequences of ignoring gender and race. *College Teaching, 46*(3), 82-88.

SCANS (1991). *What work requires of schools: A SCANS report for America 2000*. Washington, DC: U.S. Department of Labor.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition. In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 189-215). Hillsdale, NJ: Lawrence Erlbaum Associates.

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13*, 89-99.

Webb, N., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 841-873). New York: Simon & Schuster Macmillan.

White, B., Shimoda, T., & Frederiksen, J. (2000). Facilitating Students' Inquiry Learning and Metacognitive Development Through Modifiable Software Advisers. In S. Lajoie (Ed.), *Computers as Cognitive Tools, Volume Two: No More Walls* (pp. 97-132). Mahwah, NJ: Erlbaum.

## Acknowledgments

The work reported here was partially supported by The National Science Foundation (IIS/REC 0735440) and by the Edward Frymoyer Chaired Professorship.

# First-Year Engineering Students' Environmental Awareness and Conceptual Understanding with Participatory Game Design as Knowledge Elicitation

Melissa Dyehouse, Nicole Weber, Jun Fang, Constance Harris, Annette Tomory, Johannes Strobel,
Purdue University, 701 West Stadium Ave., West Lafayette, IN 47907
Email: mdyehous@purdue.edu, nrweber@purdue.edu, jfang@purdue.edu, harris11@purdue.edu, annette@golden-tech.com, jstrobel@purdue.edu

**Abstract:** The purpose of this paper was to investigate first-year undergraduate engineering students' awareness and conceptual understanding of environmental issues and to examine how effective a participatory game design strategy was to elicit their understanding. Respondents ($n$=1,394) completed baseline environmental awareness surveys and 24 respondents participated in the game design process consisting of four workshops and online activities. The game design component was focused on life cycle analysis (LCA) including environmental impacts of engineering design. Observations and artifacts were collected from workshops and interviews were conducted. Results showed that students had a general awareness about environmental issues but lacked awareness about LCA, pollution, and wetlands. Additionally, the participatory game design process showed that students struggled with applying the newly gained understandings of LCA in the game designs. The participatory game design provided a wealth of information on students' understanding and also served as an effective platform for knowledge elicitation.

## Introduction

According to the Accreditation Board for Engineering and Technology (ABET) criteria for accrediting engineering programs, engineering graduates must have the "broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context" (ABET, 2009; p. 2). One implication of this is that engineers need to be able to recognize the importance of environmental and ecological issues in engineering and to gain an understanding of what this means to them as an engineer. For example, life-cycle analysis (LCA) is a means by which to quantify the environmental impacts of a product (Blanchard & Reppe, 1998). The issues embedded within LCA are central to the design, manufacturing, use, and disposal of engineered products.

Sustainability issues play a role in all disciplines of engineering, from the consequences of the products' design to the manufacture and eventual use and disposal of the product. For this reason sustainability concepts need to be incorporated into all engineering disciplines (Boyle, 2004). Apart from programs in environmental engineering however, sustainability concepts are not well-integrated into the engineering curriculum (Crofton, 2000; Azapagic, Perdan, & Shallcross, 2005). There is also a lack of knowledge about how much first-year undergraduates in engineering programs actually know about environmental issues and how they obtained their knowledge.

The purpose of this study is to investigate first-year undergraduate engineering students' awareness of the concept of LCA and environmental issues, and examine how students' understanding can be understood within a participatory co-design of an educational game. Additionally, baseline data will provide information on (a) how influential previous educational and demographic factors are for students' environmental awareness, and (b) which environmental issues still need attention within the first year engineering undergraduate curriculum.

It is necessary that engineering students of all engineering disciplines be aware of environmental issues and that they understand the relevance of environmental issues in engineering design. For this reason, our study focused on the participatory design process as a means of increasing students' awareness and understanding of environmental issues particularly in regards to the engineering design process as well as the assessment of its effectiveness as a knowledge elicitation strategy.

*Research questions*

The core research was guided by the following questions:

- What is the environmental issues awareness of first-year engineering students?
- What are the predictors of students' environmental awareness?
- How does the game design process enhance students' awareness and understanding of life cycle assessment and sustainable engineering?

The secondary research was guided by the question: How effective is the participatory game design as a knowledge elicitation strategy to shed more light on students' understanding of environmental and sustainability issues with engineering?

## Literature and Theoretical Framework

*Participatory Design*

Participatory design (PD) was deployed in this study as a design and knowledge elicitation feature. Rooted in the Scandinavian workplace movement, "participatory design of computer applications is about the direct participation of those who will be affected by the development of a particular computer application in the decision-making, design and/or development process" (Törpel, 2005, p. 177). PD is particularly focused on the exchange of skills, values and perspectives during the development on potential working settings (Greenbaum & Kyng, 1991; Kensing & Blomberg, 1998). In this study, students are collaborative co-designers of an educational game from which other students can both give feedback in the design and benefit from the process.

The concept of PD, used primarily in product design, has long roots in educational design (Strobel, 2006) and research practice as well: Jonassen, Wilson, Wang, & Grabinger (1993) reported that instructional designers learned far more by designing CAI (computer-assisted instruction) than the target audience will probably ever learn by using the designed CAI. Additionally, the rich teach-back literature (Johnson & Johnson, 1987) shows learners are especially successful when teaching newly acquired knowledge and skills to other learners. Nevertheless, there are few research studies and reports discussing how participatory design helps shape the educational process, benefiting students when engaged as designers and end-users. In addition, the authors are only aware of a few studies utilizing participatory design as a strategy for knowledge elicitation (e.g., Kafai, Franke, Ching, & Shih, 1998).

*Environmental Education and Stewardship*

Public education on environmental issues is essential, as public opinion plays a major role in the development and implementation of environmental policy. Popular perceptions and misconceptions surrounding environmental problems must be understood in order to make educational efforts most effective (Holl, Dailey, and Ehrlich 1995). According to Holl et al. (1999), educational effectiveness requires teaching the links between different issues, the relationship between individual actions and environmental quality, and tailoring the information on these environmental problems and solutions to a more local or individual context.

Ecological literacy and more broadly environmental literacy consists of three interrelated components: knowledge, affect, and behavior (Bruyere, 2008). That is, one must know about ecology, show concern for the natural environment, and act in a way that is consistent with this knowledge and concern. It follows then, that for students to have sustainable engineering literacy they must have knowledge about the subject, show concern for the natural environment, and show behavior consistent with this knowledge and concern, for example by developing technology with the ecological footprint in mind. Several studies on environmental behavior have found links between awareness, attitudes, and behavior or behavioral intentions, suggesting that as awareness about an issue increases, so will subsequent attitudes or behaviors (Cottrell, 2003).

*Game Design as an Educational Tool*

Computer games are an emerging and popular area of development (Danielsson & Wiberg, 2006) and seen as "an important teaching tool because they can provide a compelling context via interactive, engaging and immersive activities" (Gunter et al., 2008, p. 511). Despite the hype, whether games actually help learning remains a controversial topic (Gunter et al., 2008; Rieber, 2005; Shaffer, 2006). Becker (2007) observed that "design of games for learning is one of the biggest challenges that instructional designers have had to face" (p. 43). Other research studies have shown that few games have demonstrated that they can successfully teach academic content in a classroom teaching (Gunter et al., 2008; Rieber & Noah, 1997). Frietas (2008) reported that commercial games, which focus on fun and entertainment, often lacked the principles of good learning design. Educational games, which were created as collaborative efforts between educators and instructional designers, often resulted in boring play (Lim, 2008; Prensky, 2001). Other research studies, however, began to highlight the use of students as potential educational game designers and the positive effect this experience would have on student learning (Danielsson & Wiberg, 2006; Lim, 2008). Lim (2008) argued that students would be more likely to engage in the learning process, because they would be empowered to make decisions and take action by designing or co-designing their learning experience.

## Methodology

*Research Design and Data Analysis*

In the study's research design, participants are asked to co-design an educational game with the researchers in a process called participatory design, with both quantitative and qualitative research methods employed. Quantitative methods were used to measure baseline data about participants' knowledge and understanding of environmental and ecological issues. The baseline survey was combined with academic data obtained from the University's Registrar Offices (i.e., semesters of high school math courses) and demographic data (permanent residence zip code). Students' permanent zip code data were matched to average household income data in the U.S. Census bureau from the most recent census year (2000). Qualitative methods were used as a formative measure of participants' changes in understanding through the game design process. To analyze qualitative data, two raters used the method of content and constant comparison analysis (Glaser & Strauss, 1967) to group common themes and reveal discrepancies among participants' responses.

*Procedures and Project Implementation*

Participants were 1,394 first-year engineering students who were enrolled in the ENGR100 course and who completed and returned the initial baseline survey. The sample consisted of 1,109 males and 285 females. Of those students, 1,250 were U.S. citizens and 144 were international students. The survey was administered in August 2008 and encompassed student demographic information, as students' initial understanding of ecological and environmental engineering (return rate = 82%).

For the game design component of the study, all of the ENGR100 students that participated in this survey were sent an email recruitment letter in October 2008. For the purpose of avoiding bias and increasing reliability and validity in the data collection process, participants were selected based on a stratified sampling strategy, so each participant team had a mix of demographic background. Twenty-four students participated in the game design component, developed by the research team and consisted of four workshops, an online wikispace environment, and two online activities (Figure 1). The game design workshops provided a venue in which study participants could interact amongst themselves and be observed by research team members, the wikispaces facilitated team member communication, and the online activities supplemented information presented in the workshops. Research team members also used the workshops to develop observation data, collect team artifacts, such as drawings, reflections, and provide teams with assistance when needed.

The themes of the workshops varied within the game design process. The first workshop lasted approximately one hour, and focused on Life Cycle Assessment (LCA) content knowledge, including introductions/assessment (10-15 minutes), a presentation by LCA content experts (30-40 m.), and a group activity (10-15 m.). The second workshop lasted approximately 70 minutes and consisted of a review of LCA (5-10 m.), a presentation of game design concepts (30-45 m.), and small group discussions (20-30 m.). The design of the third and fourth workshops were structured as "working sessions." Each team used their workshop time to develop game prototypes and provide one another with peer evaluation.

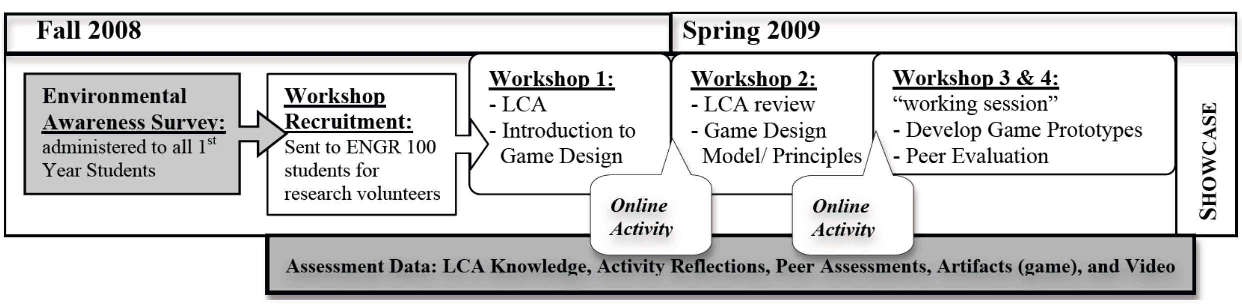



Figure 1. Workshop Activities and Timeline

*Data Collection*

Prior to the game design workshops, the research team administered a baseline survey, which consisted of several components: (a) demographic information, (b) an environmental survey (Azapagic et al., 2005) assessing students' level of awareness and understanding of sustainable development in general and about specific environmental topics, consisting of four subscales: environmental issues (14 items), environmental legislation, policy and standards (7 items), environmental tools, technologies and approaches (12 items), and sustainable development (13 items), (c) asking students to rate the importance of sustainable development, and (d) a survey about environmental knowledge from the National Environmental Education & Training Foundation (NEETF)/Roper Questions, 1997-2000 (Coyle, 2005).

Data collection for the game design process was centered on a series of game design workshops, which consisted of four face-to-face workshops and two online activities. Six teams of four students participated in the workshops and the following instruments were utilized to collect data and to triangulate the results of qualitative analyses.

- *Check-In Assessment:* Participants' reflection on their basic understanding of LCA concepts.
- *Activity Handouts:* Students reflected on activities, such as discussions and brainstorming that occurred during the workshop.
- *Artifacts:* The artifacts include the game prototypes and instructions.
- *Peer Assessments:* Each group evaluated the game paper prototypes developed by the other teams based upon a rubric developed by the research team.
- *Audio:* Includes discussion of workshop activities and final interview data.
- *Video Data:* These videos contain research team debriefing discussions held upon workshop completion.

## Quantitative Results

In the results, we will first look at the findings of the baseline survey, describing specifics of the following: self reported perspectives concerning sustainable development and environmental awareness, cultural and previous research comparisons, and predictors of environmental knowledge. Then we will discuss the effects of the game design process on environmental and life cycle analysis understanding, and the way that students integrated environmental knowledge into the game design itself.

In the baseline survey, one of the questions asked students to rate the importance of sustainable development for themselves, as an engineer, for the country, for society, for population growth, and for future generations. The percentage of students who responded "important" or "very important" was calculated for each of the categories (Table 1).

Table 1: Percentage of students' ratings of the importance of sustainable development for each category

| Category | Percentage of "Important" or "Very Important" Ratings |
|---|---|
| Future generations | 68.5% |
| Your country | 68.2% |
| Population growth | 67.1% |
| You as an engineer | 66.2% |
| The society world-wide | 65.8% |
| You personally | 61.8% |

One finding was that that approximately 35% of students did not rate sustainable development as important for any category. This indicates that engineering students may not fully grasp the importance or applicability of sustainable development in engineering or the broader society.

*Environmental Awareness*

Within the baseline survey, 12 multiple-choice items were analyzed to determine students' awareness of environmental issues and to find any gaps that may need attention within the curriculum. Table 2 displays the percentage of students who responded correctly for each question. It appears that students are most familiar with the definition of biodiversity because 92.5% of students answered this item correctly. Five other items with percentages of correct responses in the 80-89% range indicates that students are mostly familiar with the concepts assessed in these items as well, which include electricity, renewable resources, ozone, household waste, and extinction. The item that students had the least knowledge about was "What is the most common cause of pollution for streams, rivers, and oceans," with only 30.3% of students answering correctly. In addition, only half of the students responded correctly for items asking about the source of carbon monoxide pollution and the benefit of wetlands (55.2% and 55.7%, respectively). It is puzzling that environmental knowledge is particularly low when it comes to the impact of everyday behavior of end users (question 2 and 4). This knowledge assessment indicates that although students may be familiar with or have a good understanding of some environmental concepts, that there are still some important concepts they have little knowledge about.

Table 2: Percentage of students responding correctly for each item

| Item | Correct Response | % Correct Responses |
|---|---|---|

| | | |
|---|---|---|
| 1. There are many different kinds of animals and plants, and they live in many different types of environments. What is the word used to describe this idea? | Biodiversity | 92.5% |
| 2. Carbon monoxide is a major contributor to air pollution in the U.S. Which of the following is the biggest source of carbon monoxide? | Motor vehicles | 55.2% |
| 3. How is most of the electricity in the U.S. generated? | Burning oil, coal, and wood | 80.9% |
| 4. What is the most common cause of pollution of streams, rivers, and oceans? | Surface water running off yards, city streets, paved lots, and farm fields | 30.3% |
| 5. Which of the following is a renewable resource? | Trees | 82.1% |
| 6. Ozone forms a protective layer in the earth's upper atmosphere. What does ozone protect us from? | Harmful, cancer-causing sunlight | 85.4% |
| 7. Where does most of the garbage in the U.S. end up? | Landfills | 78.4% |
| 8. What is the name of the primary federal agency that works to protect the environment? | Environmental Protection Agency (the EPA) | 79.9% |
| 9. Which of the following household wastes is considered hazardous waste? | Batteries | 84.1% |
| 10. What is the most common reason that an animal species becomes extinct? | Their habitats are being destroyed by humans | 82.7% |
| 11. Scientists have not determined the best solution for disposing of nuclear waste. In the U.S., what is done with it now? | Stored and monitored | 69.2% |
| 12. What is the primary benefit of wetlands? | Help clean the water before it enters lakes, streams, rivers, or oceans | 55.7% |

Next, the self-reported levels of knowledge for four environmental subscales were analyzed. Students reported the highest level of knowledge on the Environmental Issues subscale (mean = 1.73). The Environmental Issues subscale consists of general environmental issues, such as global warming and air pollution. Students reported the lowest self-reported knowledge on the Environmental Legislation, Policy and Standards subscale (mean = 0.30); the subscale consists of specific policy and legislation questions. Finally, students still scored lower than average on the Environmental Tools, Technologies and Approaches subscale and the Sustainable Development subscale (mean = 1.16 and 1.07, respectively).

*Comparison of Self-Report and Objective Knowledge*

Previous research found that "most Americans believe they know more about the environment than they actually do" (Coyle, 2005; p. vi). To test this concept, Pearson's *r* correlation coefficient was calculated for the total awareness test score and the average self-reported awareness about environmental issues score (from environmental issues subscale questions). A significant ($p < .001$) correlation was found between the two variables. The higher a students' reported knowledge about environmental issues, the higher their actual knowledge as measured by a test of environmental issues; however the correlation was low to moderate ($r = 0.202$), indicating that strength of the correlation is weak.

*Predictors of Environmental Knowledge*

To determine what variables best predict the average awareness score of environmental sustainability, a regression analysis was performed. Standard multiple regression was considered the optimal method for answering the research question because (a) the independent variables (IVs) are correlated and there are unequal numbers of cases in cells and (b) regression can use several continuous and dichotomous IVs to represent the best prediction of the dependent variable (DV). Hypothesized variables thought to predict environmental knowledge included gender, income, previous environmental education, environmental issues subscale score, environmental policy/legislation subscale score, environmental tools/applications subscale score, sustainable development subscale score, number of semesters high school math, and number of semesters high school science.

Two variables were recoded into binary variables for better interpretation: income (low/high) and previous environmental education (none/some). The variable national/international students was omitted due to the low number of international students ($n$=144) as compared to national students ($n$=1,250). Additionally, the variable for income, although included in the model, only includes national students. Only four of the independent variables contributed significantly to prediction of environmental knowledge. These significant predictors are: (a) average self-reported knowledge of environmental issues ($\beta$=0.22, $p < .001$), (b) average self-reported knowledge of environmental tools/applications ($\beta$=0.14, $p < .01$), (c) gender ($\beta$=0.16, $p < .001$), and (d) number of semesters of HS math classes ($\beta$=0.08, $p < .01$), where $\beta$ indicates the strength of the relationship between the dependent and independent variable.

## Qualitative Results

*Students' Environmental Understanding in the Context of Participatory Game Design*

To determine what environmental awareness students gained while participating in the game design process, a content analysis was performed on the qualitative data, particularly the design artifacts, workshop observations, and follow-up interviews. The main finding that emerged was that all of the respondents ($n$ = 14) appeared to have an understanding of what LCA is, the uses of LCA, and the types of factors to consider in LCA calculations when responding to open-ended questions. It appeared that students understood the complexity of LCA and that many variables must be considered in calculating the environmental impacts of a product. For example, the most common theme that was found in the analysis of open-ended questions was *consideration of the different factors involved in LCA*. For example, typical comments were, "Deciding which factors are most/least important" and "…they [LCAs] are multifaceted [and] there is not one factor alone that determines whether one product or method is better." In contrast, all of the observers noted that of the participant teams, little to no knowledge of LCA was demonstrated at the beginning of the workshop.

At the end of the workshop, participants were asked to provide their main understanding of LCA. Out of the ten respondents, all provided either a general descriptive definition or an explicit definition of LCA. The following are examples of the two types of descriptions that participants provided.

- *Explicit (n* = 7*)*. A definition was considered explicit if a student specifically defined LCA. An example of a participant's explicit definition is, "The effect manufacturing of products has on the environment, and how to change the environmental impact of products by redesigning them."
- *General descriptive (n* = 3*)*. A definition was considered general descriptive if a student described some uses of LCA but did not specifically define the term. An example of a participant's general descriptive definition is, "My main understanding is that it is to show how you can save energy by using and building more energy proficient items."

*Integration of Environmental Information into the Game Design*

Another focus of the qualitative analysis was to determine which teams embedded to what extent LCA information into the game design. The results show that only one of the three teams explicitly embedded LCA information into the game design while the other two games had a vague reference to "improving the environment."

For example, team 4 was one of the teams that did not integrate LCA information within the game. The only decision the player must make is whether to build/expand a business or take a mission card, which could possibly be a response to the LCA criteria embedded. The mission cards allow the players to help the environment. However, this is unclear; for this team, content dissemination in the game was not consistent with the comments made by the individual group members. For example, team members made comments regarding having the player choose different options that have different LCA emphases the game reflects more about luck than choices.

Another example of a team that did not integrate LCA into the game design is team 1. In this case, players are supposed to make choices for their character with the goal of "saving the environment" and "choosing the better option," yet there are only two statements regarding environmental issues: "Player will be presented with an option based on their role…the player has to choose between improving his/ her business or saving the environment. Each decision has pros and cons that the player must weigh before choosing the 'better' option." Because Team 1 does not have data regarding personal definitions of LCA it is not clear whether the situation is similar to that of team 4, where the method of content dissemination in the game does not match what the individuals in the team articulated. The incorporation of LCA into the game became difficult to assess, and should be kept in mind for future research.

In contrast, team 2 embedded the LCA thought process into the game while developing the game, merging the concepts well. Team 2 members discussed how the game would incorporate decisions that must be made regarding the raw materials found in the environment. Although participants used the vague term "eco-index" it appears that they were thinking about the principles of LCA, as they determined the process of the game by discussing how the player will need to make choices regarding the building of the community while considering the environmental impact such as raw materials and "eco- index."

One of the reasons that students may have refrained from adding LCA into their game design was because they found the complexity of the LCA process difficult to incorporate into a game. Although a major theme that emerged from the students' written responses was the importance of taking into account the wide variety of factors in calculating LCA, this also appeared to be one of the main difficulties that students expressed. For example, several students listed the most difficult knowledge of LCA as the need to take into account many different factors and environmental impacts. For example, typical student responses to an open-ended question asking about the most difficult aspects of LCA are: "Considering the wide range of impacts that a product can have," and "…[taking] into account all that affects the environment."

## Conclusions and Implications

The purpose of this study was to investigate first-year engineering students' awareness of environmental issues and to examine how students' understanding of Life Cycle Analysis changed after participating in the process of co-designing an educational game. Results revealed that students show a lack of awareness about several aspects of environmental issues, and that all of the students who responded to the open-ended questions gained an understanding of LCA and the majority were able to define the term accurately and in detail.

Although participants responded correctly to the majority of environmental awareness questions, there were still some issues where students exhibited misunderstandings. For example, only 30% of students knew the most common cause of pollution of streams, rivers, and oceans. On a national survey, a majority of the sample of Americans also answered this question incorrectly (Coyle, 2005), indicating that there is a need to promote greater awareness about water pollution. Additional areas of concern are items relating to carbon monoxide pollution, benefits of wetlands, and nuclear waste disposal, as these issues might have implications for distributing or manufacturing certain products, and engineers should develop an awareness and understanding about these issues.

Additional results showed that there are four predictors of environmental awareness: (1) the level of awareness about environmental issues, (2) the level of awareness about environmental tools, technologies and approaches, (3) gender, and (4) the number of semesters of high school math courses. Not surprisingly, the largest predictor of environmental awareness is a students' self-reported awareness of environmental issues. Next, gender was the second largest predictor of environmental knowledge, followed by students' self-reported knowledge on the environmental tools, technologies and approaches subscale. Finally, the smallest predictor was the number of high school math courses taken, which was unexpected given that the number of semesters of high school science courses taken was not a significant predictor of environmental awareness. Thus, there may be a variable correlated with the number of semesters of high school math that was not measured and serves to predict students' environmental awareness. In addition, future research with non-U.S. citizens should use additional measures of environmental awareness because the survey used in this study may not be appropriate for international students. Lastly, students reported the lowest levels of awareness on the environmental legislation, policy and standards subscale especially as compared to the more general Environmental Issues subscale.

This study also examined how students' environmental awareness and understanding of the LCA concept changed following the game design process. Although individual participants demonstrated an understanding of the factors involved in LCA, they often did not fully incorporate their understanding into the game design. Only one of three teams explicitly embedded LCA information into the game design. Although participants showed an understanding of environmental issues as they pertain to engineering, they may have perceived LCA as too complex or lacked the time to successfully incorporate LCA into the game, indicating that additional planning time for game development is needed.

These results contribute to the research on first-year engineering students' environmental awareness. The participatory game design provided rich data about students' understanding of LCA and served as an effective platform for knowledge elicitation. It is apparent that first-year undergraduates in engineering lack awareness about many environmental issues. Because of the increased importance of sustainability in engineering, the need for engineers in all disciplines to understand sustainability issues in engineering is necessary. The game design process was a method that warrants additional research to determine the effects on students' understanding of LCA and the importance of sustainability in engineering. Future research is needed to focus more on the group processes within teams as well as how to moderate the game design process so as to better lead to participants' understanding of environmental issues in engineering and an educational game that can be used to teach others.

## References


Accreditation Board for Engineering and Technology (ABET). (2007). *Criteria for accrediting engineering programs: Effective for evaluations during the 2009-10 accreditation cycle*. Retrieved Oct. 22, 2009, from http://www.abet.org/forms.shtml

Azapagic, A., Perdan, S., & Shallcross, D. (2005). How much do engineering students know about sustainable development? The findings of an international survey and possible implications for the engineering curriculum. *European Journal of Engineering Education, 30(1)*, 1-19.

Becker, K. (2007). Pedagogy in commercial video games. In D. Gibson, C., Aldrich, & M. Prensky (Eds.), *Games and simulations in online learning: Research and development frameworks* (pp. 21-47). Hershey, PA: Information Science Publishing.

Blanchard, S., & Reppe, P. (1998). *Life cycle analysis of a residential home in Michigan* (Rep. No. 1998-5). University of Michigan: Center for Sustainable Systems.

Boyle, C.A. (2004). Considerations on educating engineers in sustainability. *International Journal of Sustainability in Higher Education 5(2)*, 147-155.

Bruyere, B.L. (2008). The effect of environmental education on the ecological literacy of first-year college students. *Journal of Natural Resources and Life Sciences Education, 37*, 20-26.

Coyle, K. (2005). *Environmental literacy in America: What ten years of NEEFT/Roper research and related studies say about environmental literacy in the U.S.* The National Environmental Education & Training Foundation, Washington, DC.

Cottrell, S.P. (2003). Influence of sociodemographics and environmental attitudes on general responsible environmental behavior among recreational boaters. *Environment and Behavior, 35*, 347-375.

Crofton, F.S. (2000). Educating for sustainability: Opportunities in undergraduate engineering. *Journal of Cleaner Production, 8*, 397-405.

Danielsson, K., & Wiberg, C. (2006). Participatory design of learning media: Designing educational computer games with and for teenagers. *Interactive Technology & Smart Education, 3(4)*, 275-292.

Freitas, S.D. (2008). *Learning in immersive worlds: A review of game based learning*. Joint Information Systems Committee (JISC), London. Retrieved Oct. 27, 2009, from http://www.jisc.ac.uk/media/documents/programmes/elearning_innovation/

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing.

Greenbaum, J., & Kyng, M. (Eds.). (1991). *Design at work: Cooperative design of computer systems*. Hillsdale, NJ: Lawrence Erlbaum.

Greenbaum, J., & Stuedahl, D. (2000). *Deadlines and work practices in new media development*. Proceedings of IRIS 23, Laboratorium for Interaction Technology, University of Trollhattan Uddevalla, 2000.

Gunter, G.A., Kenny, R.F., & Vick, E.H. (2008). Taking educational games seriously: Using the RETAIN model to design endogemous fantasy into standalone educational games. *Education Technology Research and Development, 56*, 511-537.

Holl, K.D., Daily, G.C., Daily, S., & Ehrlich, P.R. (1999). Knowledge and attitudes toward population growth and the environment: University students in Costa Rica and the U.S. *Environmental Conservation, 26,* 66-74.

Holl, K.D., Daily, G.C., & Ehrlich, P.R. (1995). Knowledge and perceptions in Costa Rica regarding environment, population, and biodiversity issues. *Conservation Biology*, *9*, 1148-1558.

Johnson L. & Johnson N.E. (1987). Knowledge elicitation involving teachback interviewing. In A. Kidd (Ed.), *Knowledge elicitation for expert systems: A practical handbook* (pp. 91-108). New York: Plenum Press.

Jonassen, D.H., Wilson, B.G., Wang, S., & Grabinger, R.S. (1993). Constructivistic uses of expert systems to support learning. *Journal of Computer Based Instruction, 20*, 86-94.

Kafai, Y.B., Franke, M.L., Ching, C.C., & Shih, J.C. (1998). Game design as an interactive learning environment for fostering students' and teachers' mathematical inquiry. *International Journal of Computers for Mathematical Learning, 3*, 149-184.

Kensing, F., & Blomberg, J. (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work, 7*, 167-185.

Lim, C. (2008). Spirit of the game: Empowering students as designers in schools? *British Journal of Educational Technology, 39*, 996-1003.

Rieber, L. (2005). Multimedia learning in games, simulations, and microworlds. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*, (pp. 549-567). Cambridge: Cambridge University Press.

Rieber, L.P., & Noah, D. (1997). *Effect of gaming and graphical metaphors on reflective cognition within computer-based simulations*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Prensky, M. (2001). *Digital game-based learning*. New York: McGraw Hill.

Shaffer, D.W. (2006). *How computer games help children learn*. New York: Palgrave Macmillan.

Strobel, J. (2006) Participatory design strategies for eLearning: A design-based research approach in the field of Educational Technology. In J. Multisilta (Ed.). *Proceedings of the workshop on human centered technology*. Pori Publications, Tampere University of Technology.

Törpel, B. (2005). *Participatory design: A multi-voiced effort*. 4th Decennial Conference on Critical Computing: Between Sense and Sensibility, Proceedings, 21-25, Aarhus, Denmark, pp. 177-181.

# "I Don't Know—I'm Just Genius!": Distinguishing Between the Process and the Product of Student Algebraic Reasoning

Jose Francisco Gutierrez, Graduate School of Education,
University of California—Berkeley, Berkeley, CA 94720-1670, josefrancisco@berkeley.edu

**Abstract:** This paper reports on a semiotic-cultural analysis of two high-school minority seniors' participation in a mathematics instructional intervention conducted at a school for academically at-risk students. Focusing on the students' speech acts, gestures, and artifact production during their successful collaborative engagement in an algebraic pattern-finding task, I adopt and adapt Radford's (2003) framework to evaluate the quality of each student's actual learning as implicated by their discursive contributions. By analyzing each student's utterances as either "to" a semiotic mode (generalization) or just "in" the mode, and then overlaying these coded utterances, I show how one student's participation, which appears to mark a developing algebraic command, in fact blinds the teachers to underlying discontinuity in the student's meaning construction. The study illuminates critical tradeoffs in the design and facilitation of collaborative problem solving.

## Introduction: The Story of Be'zhawn and Deana

We join Be'zhawn and Deana, two high school seniors classified by their mathematics teacher as low-achieving students, in the midst of a tutoring session involving a collaborative algebraic problem-solving activity. Be'zhawn and Deana are at the board, working together on a sequence of several geometric figures in an attempt to formulate its pattern algebraically (see Figure 1 below). For twenty-five minutes, the students have been bandying mathematical ideas about: some ideas are taken up and built upon, while others are ignored or proven inefficacious. Throughout this duration, both students sustain a high level of engagement and, to all appearances, contribute equally to the problem-solving process. Eventually, Be'zhawn is the first to verbally express a correct solution. And yet, when asked to explain his reasoning, he replies, "I don't know—I'm just genius!"

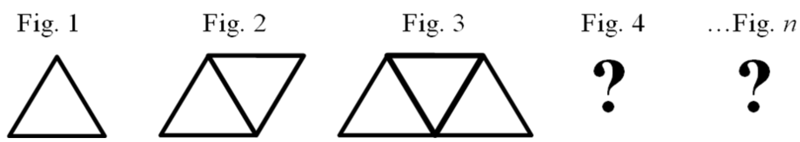



Figure 1. The triangle sequence—a paradigmatic reform-based algebra problem. The task objective is to express $U_n$, the total number of line segments in the $n^{th}$ figure. For example, "Fig. 1" consists of 3 line segments, "Fig. 2" consists of 5, "Fig. 3" consists of 7, etc., so that the $n^{th}$ figure consists of $2n+1$ line segments.

Back in our laboratory, all three teachers—the author included—discussed the students' behaviors during this activity and the nature of their mathematical contributions. We all agreed that this particular tutoring session was unique, as compared to previous sessions, in terms of the quality of students' mathematical reasoning. Also, as this was the first time they had collaborated on a pattern-finding problem, we attributed the students' engagement to the problem type's unique task structure. Be'zhawn, in particular, had exhibited much difficulty with algebraic concepts in prior sessions, where the problem items had been more routine. The pattern-finding problem was not the only item Be'zhawn solved quicker than his peers; however, this was the first time he had ever indicated that he held a positive perception of himself as a mathematics doer, when he confidently declared that he was "genius."

This particular tutoring episode was part of a larger mathematics-education research project conducted at a small alternative high school in the San Francisco Bay area. Broadly, we sought to identify and respond to learning challenges faced by academically at-risk mathematics students with respect to both reform and traditional algebra curricula. The pattern-finding episode appeared to demonstrate success and not only challenge, and so I selected this episode in an attempt to articulate principles of design and facilitation that contributed to its apparent success. That said, my enthusiasm was tempered by acknowledging that Be'zhawn's inability to explain his correct solution might implicate his solution strategies as problematic. What might it mean for a student to solve an algebraic pattern-finding problem yet not be able to explain his solution? What methodological tools could address this puzzle?

## Objectives of the Study

Researchers of mathematics education constantly develop theoretical frameworks for articulating what students struggle with as they attempt to solve algebraic problems. At the same time, though, the national achievement gap is constantly increasing. This, despite constant calls for equity in mathematics education and, in particular, for increasing the accessibility of algebra content for students from minority groups and economically disadvantaged backgrounds (for a review see DiME, 2007). The current study embarks from a conjecture that some recent theoretical frameworks developed particularly for tackling algebraic reasoning (for a review see Kieran, 2007) harbor the potential to illuminate new directions for broadening diverse participation in mathematics and, in particular, to support general success in this "gatekeeper" content. The objective of this paper is to build upon some delimited empirical data so as to illustrate what we may need to attend to as we pave new paths through the gate.

Central to the theoretical argument of this paper is Luis Radford's (2003, 2008) *semiotic-cultural approach* to theorizing the process and content of students' algebraic reasoning. This powerful approach views learning as an evolving process co-constrained by both cognitive and socio-cultural factors. Specifically, mathematics learning is conceptualized as constructing personal meaning for canonical semiotic artifacts (e.g., algebraic symbols such as the variable "*x*"). Through consolidation and iteration of such constructions, students appropriate these notations and, reciprocally, build personal meaning for mathematical content as well as fluency with the disciplinary procedures. That is, individuals who build mathematical signs upon personal meaning develop more than fluency with procedures—they develop substantive and articulated understanding of the targeted content. Radford's approach takes into account a vast arsenal of personal and interpersonal resources that students bring to bear in solving mathematical situations, including linguistic devices and mathematical tools. Most relevant to our current study, Radford has demonstrated that critical steps within individual learning trajectories can be explained by noting subtle shifts in the subjective function and status of the semiotic artifacts (Abrahamson, 2009; Sfard, 2007).

I here propose Radford's approach as a powerful tool for untangling the compound challenges faced by students who are attempting to develop a mathematical register even as they are learning new content. Specifically, I propose, the semiotic-cultural approach enables researchers to delve into the cognitive/semiotic plane below external discursive manifestations. That is, students may manifest discursive productions in a proper mathematical register, yet these "correct" utterances may belie the absence of underlying personal meaning. The semiotic-cultural approach, I submit, could be leveraged so as to expose tension between students' overt mathematical speech contributions and their covert, (dis)connected reasoning.

A potential theoretical contribution of this paper is the qualification I propose for Radford's approach, as I adapt the approach to the particular needs of at-risk students. These students' disposition toward the classroom "leading discourse" (Sfard, 2007) is complex, because for them, "talking the mathematical talk" is at once enabling and threatening. We therefore may witness gaps between these students' overt expression and covert meaning. This tension bore out in the analysis. Namely, Whereas Radford's approach was pivotal to making initial sense of these students' utterances, I gradually realized, however, that the approach was not enabling me to unwrap students' covert meanings. As the paper will demonstrate, I was thus impelled to revisit some of the implicit assumptions of Radford's semiotic-cultural framework and then make a pivotal amendment. Namely, whereas Radford's interpretation of collaborative problem solving may imply that all team members are party to critical semiotic transitions inherent to generalization acts, I submit that some students manifest productions *in* milestones of this procedure without having partook in advancing *to* these semiotic modes. Distinguishing the within-vs.-between semiotic nature of students' discursive productions is important, because, as Radford claims, it is the *transitions* from one semiotic mode to the next—not discursive acts per se—that are crucial to building meaning (Radford, 2003). Equipped with this qualification, the semiotic-cultural approach helps me characterize weaknesses in some students' learning processes during collaborative problem solving as resulting from an impetus to participate within a particular semiotic mode without necessarily having personally made the careful transition to that mode.

## Theoretical Background

From the semiotic-cultural perspective, mathematics learning is viewed as a mediated, distributed, and dynamically reciprocal process in which students' emerging presymbolic knowledge is iteratively objectified in progressively sophisticated forms of historically and culturally endowed semiotic systems. One such semiotic system—which is prevalent in the mainstream of school mathematics programs—is algebra. Viewed from the semiotic-cultural perspective, "algebraic thinking is a particular form of reflecting mathematically" (Radford, 2003), that is, algebra is about "using signs to think in a distinctive way" (Radford, 2008, p. 87).

To illuminate the content and process of this distinctive way of thinking characteristic of algebra, Radford elaborates a theoretical framework for describing and analyzing students' reasoning during collaborative engagement with pattern-generalization algebraic activities such as the triangle sequence (see Figure 1, above). I now use this problem as a context to elaborate on Radford's framework, which is central to my data analysis, and then outline a potential shortcoming with its extant formulation and finally propose a solution to this problem.

A key construct in Radford's framework is *knowledge objectification*, which is defined as the process of making the objects of knowledge apparent (Radford, 2003). For example, a mathematics learner, in an attempt to convey a certain aspect of a concrete object, such as its shape or size, will tap a variety of semiotic artifacts such as mathematical symbols and inscriptions, words, gestures, calculators, and so forth. In patterning activities, however, some of the objects of knowledge are general and therefore "cannot be fully exhibited in the concrete world" (Radford, 2008, p. 87). More generally, mathematical notions may not be cognitively accessible, because they do not exist in the world for empirical investigation (Duval, 2006),that is, these notions are never apparent to direct perception. Therefore, in order to instantiate (objectify) these ephemeral notions, students must resort instead to personally and culturally available forms such as linguistic, diagrammatic, symbolic, and substantive artifacts as well as the body, which Radford (2003; 2008) collectively terms *semiotic means of objectification*.

In patterning activities, students often resort to using one of two strategies when they attempt to construct generalizations from an initial set of numeric and/or figural cues (Radford, 2008) such as in the triangles sequence.

1. The first strategy, *naïve induction* is characterized by a process of disconnected or minimally principled "guess and check." For example, a student might propose a simple rule such as "the figure-number times two" and check its validity on a few cases. Finding that the proposed rule does not satisfy one or more cases, the student might then modify her rule or propose an entirely different rule, for example, "the figure-number times two, plus one" and check its validity as well, and so on.
2. The second and more sophisticated strategy, *generalizing*, is expressed as an active search for recurrent relationship structures and/or pattern commonalities among constituent elements in the problem space. This latter strategy, in turn, typically produces two types of generalization, *arithmetic* and/or *algebraic*.
   2.1. Arithmetic generalizations typically take the form of a *recursive solution* that only indirectly expresses any term in the sequence. In the triangle sequence, the number of line segments required to construct each consecutive figure always increases by a factor of two with respect to the previous figure, an observation expressible as the arithmetic generalization $U_n = U_{n-1} + 2$. The commensurate *explicit solution* $U_n = 2n + 1$, however, is more powerful, because it determines directly any item along the infinite sequence.
   2.2. Making the leap from recursive to explicit solutions involves generalizing a pattern *algebraically*, which requires grasping and objectifying recurrent $n$-to-$U_n$ relations along the sequence, then elaborating this objectification into a generalization that would *directly* express any term in the projected sequence.
   2.3. Finally, algebraic generalizations are subdivided into three types in accordance with their level of generality: *factual, contextual,* and *symbolic* (see Figure 2, below, for a summary as well as examples).

| | | | | | |
|---|---|---|---|---|---|
| **Solution Strategies** | Naïve Induction<br>• Characterized by "guess and check"<br>• E.g., "It's '2*n*+2'… No. Try '2*n*+1'. YES!" | Generalizing<br>• Active search for recurrent relationships/commonalities among constituent elements in the problem space | | | |
| **Generali-zation Types** | | Arithmetic Generalization<br>• Recursive (*U* is expressed in terms of previous figure)<br>• E.g., $U_n = U_{n-1} + 2$ | Algebraic Generalization<br>• Explicit formula (*U* is expressed in terms of the figure's ordinal position)<br>• E.g., $U_n = 2n + 1$ | | |
| **Levels of Generality** | | | Factual<br>• Objects and operations are bound to concrete level<br>• E.g., "1 plus 2, 2 plus 3, 3 plus 4…" | Contextual<br>• Objects and operations are abstracted and generalized<br>• E.g., "The figure plus the next figure." | Symbolic<br>• Objects and operations are expressed through formal symbolism<br>• E.g., "$n + (n+1)$" |

Figure 2. Semiotic-cultural taxonomy of algebraic pattern-finding solution processes (adapted, Radford, 2003; 2008)

## A Proposed Qualification

When I first applied the semiotic-cultural framework to the data (see Analysis Techniques, below), it indeed revealed disconnection between the students' utterances and their underlying meaning making. Namely, Deana's utterances reflected generalizing acts, whereas Be'zhawn's did not. Essentially, applying Radford's framework revealed that one student 'got it' while the other did not 'get it.' However, what the extant framework did not explain was why it appeared to the instructors in situ that *both* students 'got it.'

A second analysis of the data was conducted so as to provide an alternative description of the students' behaviors—a description that would explain how the researcher-as-analyst could see what the researcher-as-teacher could not. So doing, I realized that inherent to Radford's framework were what might be called three *semiotic modes* of action, in which both Deana and Be'zhawn were participating. By "mode" I am offering an analytic construct that is linguistic but not psychological; that is, by "mode" I am referring to the surface features of student utterances, temporarily putting aside the question of the student's subjective semiotic grounding. I could now conjecture that students' utterances *within* each of these semiotic modes were similar, thus giving the teachers the impression that both students shared similar learning experiences toward those utterance.

By introducing "modes" as an elaboration to the framework, I differentiate two meanings of "generalization": the *process* of generalizing (i.e., the way that students engage with the problem), and the *product* that results from this process. This distinction is important, because it suggests that researchers and teachers may attribute understanding to students when such attribution is not warranted. In particular, a student's utterance *within* a semiotic mode may be interpreted as marking the production of a semiotic generalization *to* that mode, when in fact the student is only appropriating and elaborating on the product of another student's generalization to the mode.

The proposed process-vs.-product distinction reflects a pedagogical position that the generalizing activity is not merely to create opportunities for students to operate *in* each of the modes, but rather to support them as they progress *along* the desired chain of signification from the factual through to the symbolic mode, thus ensuring that each and every student performs the cognitive work necessary to personally ground the target signs.

Having presented a semiotic–cultural framework and outlined a potential problem with its extant formulation as well as a proposed solution to this problem, I am now in a position to restate my research questions in light of Radford's seminal theory. Namely:

- Can a mathematics learner be *in* a particular semiotic mode without having generalized *to* that mode?
- What are the consequences of such superficial semiotic production for learning?

Next I describe the methods used to address these questions.

## Methods

For this preliminary empirical study, the research team, which included the author as well as two undergraduate apprentice researchers, conducted an instructional intervention—*The Mathematics Workshop*—at a small high-school for academically at-risk students. The intervention, which was designed to prepare a small cohort of senior students (N=5) for a state exit exam, provided one-to-one and small-group tutoring sessions with a focus on algebra content. The research team all shared in the design, facilitation, and preliminary study of the workshop. For this paper I focus on two case studies, Be'zhawn (male) and Deana (female; both pseudonyms), and conduct my data analysis from the semiotic-cultural perspective. To address my research questions (see above), I examined students' speech acts, gestures, and artifact production during collaborative engagement with the "triangle sequence." I analyzed each of these students' reasoning processes during the focus episode in terms of its semiotic trajectory and then juxtaposed these two trajectories. This juxtaposition reveals "hidden" dyad dynamics in the joint activity of interpreting the sequence and constructing its algebraic reformulation.

*Data Sources*: The entire data corpus from the intervention includes students' original work, a total of six hours of video footage from the implementation, and a project wiki (online collaborative archive) that we used to coordinate resources, document fieldnotes and meeting minutes, and share ongoing reflections. However, the study reported in this paper focuses on a span of only 28 minutes of video footage, in which the triangle sequence was implemented. I selected this particular teaching episode for deep analysis both because this activity has been previously studied (e.g., Radford, 2003; Rivera, 2008) and because the participants' behaviors as they engaged in the solution of this problem helped me first notice the process-vs.-product of semiotic constructions.

*Materials and procedures*: The teaching episode in question begins with an utterance from the first teacher (hence "T1," the author; "T2" = the second teacher, one of the undergraduate apprentice researchers; "T3" = the third teacher, the other undergraduate apprentice researcher). T1 sketched the first three figural cues on the whiteboard and instructed Deana and Be'zhawn to search for a mathematical pattern and express it as an algebraic equation. Aside from these initial figural cues, a set of guiding questions was also written on the board: "How many lines are there in Fig. 1? Fig. 2? Fig. 3? Fig. 4? Fig. 12?" Furthermore, a key design feature for implementing the triangle sequence was to substitute increasingly larger numbers (e.g., Fig. 50, and Fig. 67) as a way to impress upon students that ultimately the arithmetic/recursive strategy is inefficient, thus motivating the need for more powerful tools and strategies such as algebraic generalizing as well as the use of explicit formulas.

## Analysis Techniques

I produced and analyzed a transcription of the 28-minute teaching episode, which captures all verbal, gestural, inscriptional, and other semiotic actions that were clearly observable in the video. The transcription was parsed into a total of 440 "Lines," where each Line is defined as a segment of speaker continuous speech. Next, I identified all student utterances involving mathematical propositions (127 out of the 440 Lines). For this study, I focus only on these Lines of the transcription, for which two main analytic questions were asked pertaining to its semiotic nature:

1. Generalization Type (generalization *to* the semiotic mode—i.e., the *products*):
   1a. Is the expression the product of naive induction or generalizing?
   1b. [If so,] Is the expression an arithmetic or algebraic generalization?
   1c. [If so,] Is the expression a factual, contextual, or symbolic generalization?
2. Semiotic Mode (operating *in* the semiotic mode, and therefore having ostensibly arrived *to* the mode and ostensibly engaging in the *process* of generalizing to the next mode):
   2a. Is the student referring to and/or expressing an action upon concrete objects (i.e., operating in the factual mode)?
   2b. Is the student referring to and/or expressing an action upon abstract yet contextually situated objects (i.e., operating in the contextual mode)?
   2c. Is the student referring to and/or expressing an action upon algebraic symbolical objects (i.e., operating in the symbolic mode)?

Working with both the video footage and the abbreviated transcription, a first pass of the data was done using analytic questions 1a-1c (see above). I initially evaluated whether or not each utterance reflected a generalization *to* a particular mode or merely a reiteration *in* the mode. This evaluation was based on a qualitative microgenetic analysis (Schoenfeld, Smith, & Arcavi, 1993) of students' behaviors during their conversation. So doing, I determined whether the students engaged in authentic generalizing acts (i.e., grasping and objectifying recurrent $n$-to-$U_n$ relations and providing a direct expression for any term along the sequence) or resorted instead to naïve induction and/or appropriating the speech forms of others. Following this qualitative analysis, a second pass through the data was done using analytic questions 2a-2c, whereby all student utterances involving mathematical propositions were coded only for their semiotic mode (factual, contextual, or symbolic). For example, the student utterance "Fig. 3 has seven lines" was coded as an assertion *in* the factual mode, because Fig. 3 was perceptually available when the utterance was verbalized (i.e., Fig. 3 was drawn on the board; it was "present"). And yet, the utterance "Fig. 5 is going to have eleven" was also coded as *in* the factual mode even though Fig. 5 was not perceptually available (i.e., Fig. 5 was not drawn on the board; it was "absent"), because the utterance "Fig. 5" refers to a specific figure along the sequence, which could be instantiated as a concrete entity. In contrast, utterances such as, "We just add a triangle to every one" were coded as assertions *in* the contextual mode because the student expressed an action upon abstract objects, whereby "one" is a pronoun referring to the general term "figure." Finally, transcript Lines containing students' inscriptional actions that involve symbolical notation were coded as assertions *in* the symbolic mode. For example, the inscription "$x2 + 1$," which a study participant wrote on the board, was coded as *in* the symbolic mode.

## Results and Discussion

The results and the discussion begin by presenting a qualitative analysis of several transcription segments, followed by further quantitative analysis of students' utterances in terms of their semiotic mode. I present the results in this order—which is atypical of social science research articles—because I wish to provide for the reader a chronological description of how my initial investigation of the data led to the need for theoretical extension. Combined, these analyses reveal that the students' contributions during their conversation occupied each of the three modes, however, Deana's but not Be'zhawn's utterances generalized *to* these semiotic modes.

## Qualitative Results

I here present qualitative data analyses of a series of selected transcript segments. Due to space constraints, these selected segments do not represent all of the students' mathematical activity during the 28-minute episode; however, these segments are representative of the students' behaviors during the episode in question. With these segments, I aim to elaborate on findings from the quantitative analysis by illustrating that: (1) both Deana and Be'zhawn's discursive production were *in* each semiotic mode; and yet by-and-large (2) Deana's manifestations ultimately generalized *to* each mode, whereas Be'zhawn's did not.

*Segment 1*: At the onset of the activity, Deana immediately noticed that the figures could be construed as a succession of accruing triangles; she also noticed that the number of line segments for each consecutive figure increased by an addend of two. Deana continued to operate both *in* the factual and contextual mode. So doing, she constructed an arithmetic generalization *to* the contextual mode in the form of $U_{n+1} = U_n + 2$. Deana thus began a process of authentic generalizing *in* the factual and contextual modes and this qualitative analysis suggests that her subsequent semiotic productions could be grounded in concrete elements from the initial problem space. [Each segment of data presented below is labeled by its Line number from the original transcription, the speaker, utterance, and its ostensible semiotic mode: "F" = factual; "C" = contextual; "S" = symbolic]

| | | | |
|---|---|---|---|
| Line 14 | Deana | "[*gazes at Fig.s 1, 2, and 3 drawn on board*] Fig. 4 is gonna have nine." | F |
| Line 16 | Deana | "We just add a triangle to every one, right?" | C |
| Line 33 | Deana | "[*Fig.*] 4 has nine. It goes up every two lines, so Fig. 5 is going to have eleven…" | C |

*Segment 2*: This segment illustrates how Be'zhawn first appropriates Deana's utterance and builds upon it, such that his speech acts are *in* (but not *to*) the factual and contextual modes. Furthermore, Be'zhawn gave no indication that he had interpreted the original problem-situation and had not noticed the patterns that Deana was verbalizing.

| | | | |
|---|---|---|---|
| Line 47 | Deana | "[*addressing no one in particular, thinking aloud, referring to Fig. 12*] Twelve and two… no… twelve and two is twenty-four." | F |
| Line 49 | Be'zhawn | "Twenty-four! That's the answer. […] What's twelve times two? Eleven times two is twenty-two. Ten times ten is twenty." | F |
| Line 57 | Be'zhawn | "They all go up by two." | C |
| Line 68 | Teacher 1 | "[*after asking Be'zhawn to draw Fig. 4 on the board*] They're just triangles, right?" | -- |
| Line 69 | Be'zhawn | "Oh, they're just triangles? What the—" | -- |

*Segment 3*: For the remainder of the activity, Be'zhawn employed naïve induction. For example, he guessed that Fig. 67 is comprised of 33 lines (i.e., $U_n \approx n / 2$), and then changed his guess to 134 (i.e., $U_n \approx n*2$).

| | | | |
|---|---|---|---|
| Line 191 | Be'zhawn | "Thirty-three... That's how many lines [*Fig.*] 67 has. Nah, that's too little. [*Fig.*] 12 is twenty-six." | F |
| Line 199 | Be'zhawn | "But it's not thirty-three though, cuz only [*Fig.*] 12 is twenty-six." | F |
| Line 254 | Be'zhawn | "I think it's one hundred and thirty-four." | F |

*Segment 4*: Interestingly, it was Be'zhawn who first produced a non-symbolic version of the correct solution procedure. Assisted by the teachers, Be'zhawn noticed that his proposed rule, $U_n = n*2$, was off mark by a difference of 1. Using this piece of new information, Be'zhawn cobbled together a correct solution procedure, namely, $U_n = n*2+1$, and he used it to articulate the solution for the case of Fig. 67.

| | | | |
|---|---|---|---|
| Line 308 | Teacher 2 | [*to Be'zhawn*] That was the guess, but these are each off a little bit so what do we think this is actually gonna be? | -- |
| Line 309 | Be'zhawn | Unless you add one to it. | C |
| Line 310 | Teacher 1 | So what do you think the actual— | -- |
| Line 311 | Be'zhawn | One-thirty-five! | F |

Even though Be'zhawn's solution is mathematically correct, I argue that his semiotic constructions were disconnected from the original problem space, because they were the result either of naïve induction or of appropriating the speech forms of others, mainly Deana's. Namely, my analysis reveals that Be'zhawn was guessing and operating on proximal numerical values in isolation from the initial set of figural cues. At least, Be'zhawn could not reconstruct how he arrived at these constructions. Indeed, when a teacher asked Be'zhawn, "What's the formula?" he responded as follows: "I don't know—I'm just genius!"

*Segment 5*: Deana, on the other hand, was able to build upon her initial arithmetic generalization (see Segment 1 above). So doing, at the end of the activity she was able to connect components of the solution procedures to elements from the original problem situation. Thus, I maintain, her inscription "*x*2 + 1," which she wrote on the board, was semioitcally grounded, from the factual through to the symbolic.

| | | | |
|---|---|---|---|
| Line 366 | Deana | "The formula. Okay. You know sixteen, right? Times two, multiply by two, is thirty-two, right? "And you add one more cuz of the extra line, [*gestures the shape of the top of a triangle* "/\"] cuz it's three lines [*referring to missing bottom line*]. And you just—since it's three lines." | F |
| Line 406 | Deana | "It's gonna work because you got the two and then you add one more line." | C |
| Line 414 | Deana | "[*writes* "*x*2 + 1 = ?" *on the board*] *x* times two plus one." [*Be'zhawn walks to the other side of the board, gazes toward Deana's inscription, then he writes "*Figure x 2 + 1*," where the "*x*" is read as a multiplication symbol*] | S |

The qualitative analyses above illustrate how Deana was able to construct both arithmetic and algebraic generalizations, whereas Be'zhawn was not. We see the subtle shifts in Deana's perception of the function and status of the semiotic artifacts—beginning with her initial grasp of a recurrent relationship between figures, "we just add a triangle to every one" (Line 16), to her arithmetic generalization expressed in terms of the line segments, "it goes up every two lines" (Line 33), to her final algebraic inscription of "*x*2 + 1 = Lines" (Line 414). Be'zhawn's semiotic productions, on the other hand, certainly manifested *within* each of the generalization modes, however, his shifts *between* the different modes are weakly connected at best, and non-existent at worst.

## Quantitative Results

Figure 3, below, represents students' semiotic-mode participation sequences—it depicts a synoptic left-to-right view of each student's participation through the lens of the semiotic mode they are *in*. Every node in the graph represents a student utterance coded as articulated in one of the three semiotic modes. By overlaying their respective semiotic sequences, it may appear that both Deana and Be'zhawn are making forward advances between modes, yet a closer analysis of their manifest productions reveals that this is not the case.

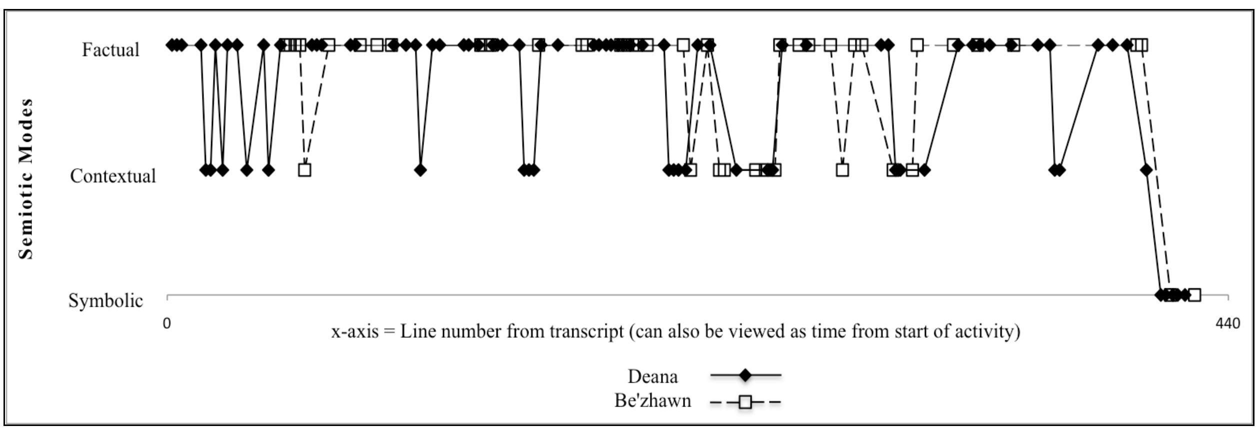



Figure 3. Representation of juxtaposed student participation sequences (ostensible semioitc trajectories)

Figure 3 clearly illustrates that Be'zhawn and Deana's respective semoitic-mode participation sequences were indeed similar, which could acccount for why the teachers, in situ, interpreted their bahaviors as indicating similar and mutually beneficial reasoning processes. And yet, in Figure 3, it is important to notice that Deanna's move from the factual to the symbolic (at the very end) occurs before Be'zhawn's move. We also see that Deanna's sequence passes through the contextual mode en route to the symbolic whereas Be'zhawn's does not, an observation supporting the hypothesis that his semiotic productions were disconnected, as apparent in his "I don't know" statement mentioned earlier in my qualitative analysis.

## Conclusions and Implications

When students participate in collaborative problem solving, their discursive productions do not necessarily reflect adequate grounding of mathematical symbols, even in the cases where these contributions are instrumental to the group's overall success. Namely, while participating in collaborative problem solving, students' individual discursive productions can occupy each of the three semiotic modes identified by Radford, yet these manifest productions might nevertheless conceal a discontinuous semiotic trajectory, resulting in suboptimal learning. Previous research has challenged the claim that mathematics learning is best done in small groups or dyads (e.g., Barron, 2003), and this paper contributes to this previous research by providing a semiotic-cultural articulation of the variability of meaning construction for individuals participating in a dyad. Notably, the paper certainly does not argue against collaboration as a mode of classroom practice. Rather, it has been my intention to provide analytic tools that problematize any assumptions that educators might harbor with respect to issues of equity that emerge when two or more students are assigned a joint problem-solving task (cf. Esmonde, 2009).

In addition to these issues of design and facilitation, the paper has also made two tentative contributions to mathematics-education research: (1) a theoretical contribution is a proposed refinement to Radford's semiotic-cultural approach; and (2) a methodological contribution is the initial development of an analytic approach to distinguishing between the process and product of students' discursive contributions to collaborative (algebraic) problem solving.

## References

Abrahamson, D. (2009). Orchestrating semiotic leaps from tacit to cultural quantitative reasoning—the case of anticipating experimental outcomes of a quasi-binomial random generator. *Cognition and Instruction, 27*(3), 175-224.

Diversity in Mathematics Education (DiME). (2007). Culture, race, power, and mathematics education. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 405–433). Charlotte, NC: Information Age.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics, 61*(1&2), 103-131.

Esmonde, I. (2009). Ideas and identities: Supporting equity in cooperative mathematics learning. *Review of Educational Research, 79*(2), 1008-1043.

Kieran, C. (2007). Learning and teaching algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 707 – 762). Charlotte, NC: Information Age.

Radford, L. (2003). Gestures, speech, and the sprouting of signs: A semiotic-cultural approach to students' types of generalization. *Mathematical Thinking and Learning, 5*(1), 37-70.

Radford, L. (2008). Iconicity and contraction: A semiotic investigation of forms of algebraic generalizations of patterns in different contexts. *ZDM: The International Journal on Mathematics Education, 40*(1), 83-96.

Schoenfeld, A. H., Smith, J., & Arcavi, A. (1993). Learning: The microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. IV, pp. 55 – 175). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sfard, A. (2007). When the rules of discourse change, but nobody tells you: Making sense of mathematics learning from a commognitive standpoint. *Journal of the Learning Sciences, 16*(4), 565-613.

## Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education pre-doctoral training grant R305B090026 to the University of California, Berkeley. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education. A very special thank you to the students who participated in this study and the school principal and staff for letting me visit their school. I also appreciate the very helpful comments provided by the three anonymous reviewers of this paper for ICLS.

# Contingent Identification in a Biomedical Engineering Classroom

Vanessa Svihla, University of California, Berkeley, vsvihla@hotmail.com

**Abstract:** This study, set in a university biomedical engineering capstone design course, explores how students identify as engineers, finding this process to be contingent and dynamically negotiated. Biomedical engineering, as a degree program, serves as a legitimate pathway towards engineering and medical school, making identification with engineering particularly contingent. Living narratives provide glimpses into experiences students have in their design course, which aligns more closely with authentic engineering practice than their previous coursework. As a result, the design course is a significant opportunity for students to consider themselves to be engineers. Framed by changes in *accountable disciplinary knowledge* and *navigation* (Stevens, O'Connor, Garrison, Jocuns, & Amos, 2008) within the interdisciplinary context of biomedical engineering and the varied goals of students for choosing the major (engineering vs. health sciences), students' identification as engineers emerges as particularly contentious. Leadership and underrepresentation are considered in relation to identification, navigation, and accountable disciplinary knowledge.

*Everything I have learned comes into play.... All I can say is that I think like an engineer. I see things from many different angles. I calculate. I am much more realistic.*

*I mean, like a surgical procedure? I mean, that's not considered any sort of technical…? Like, a surgical procedure? And, like, taking readings in a rat's stomach? I mean, something that nobody else has* ***even*** *ventured to do in* ***this*** *course?!*

## Background

These two quotes from students express tensions between two predominant and legitimate pathways (biomedical engineering and medical school) students pursue while yet positioned as "BMEs"- students in biomedical engineering. This paper explores how students working in teams in a university capstone biomedical engineering course contingently express their identification with engineering as they learn to design. Ethnographically-collected lived narratives (Ochs & Capps, 2001) provided glimpses into student experiences during this process, highlighting tensions and dimensions of contingency in students' identification as engineers, and how this intersects with their learning. These are viewed through three *dimensions* for understanding pathways through engineering education: *accountable disciplinary knowledge* (ADK) - that is, the work and practices that *count* as engineering at a given point; *navigation* of pathways; and *identification* with and by engineering (Stevens, et al., 2008). Additionally, underrepresentation and leadership are considered in terms of how they might intersect with these dimensions of ADK, navigation, and identification.

The capstone model provides a fertile ground for observing learning and related shifts in identification, taking learning to be an "experience of identity"(Wenger, 1999, p. 215). The social process of engineering design (Bucciarelli, 1994) and learning to design in a team involves the coordination of many lenses. The capstone design course introduces new identity spaces for the students: They may form a team identity as they develop a sense of belonging in a consequential manner to their team (Guzzo, 1986). They negotiate various design decisions (Bucciarelli, 1994), learning by explaining perspectives (Webb, Troper, & Fall, 1995) and by comparing to team mates' perspectives (Phelps & Damon, 1989). Within a team, each member assumes roles such that in total, all steps are accomplished, but not by all team members (Goldschmidt, 1996). Gunther, Frankenburger, and Auer (1996) caution that though enacted roles have been noted, understanding how and why roles form is largely unexplored for design contexts. By exploring students' identifications as engineers, we may shift focus to understanding some of the dimensions students face along their paths towards or away from engineering. Additionally, understanding how individuals from underrepresented groups identify as engineers may elucidate some of the experiences that contribute to their (dis)engagement (Heyman, Martyna, & Bhatia, 2002; Seymour & Hewitt, 2000).

## Setting and Participants

The participants of this study were undergraduate senior biomedical engineering students enrolled in the capstone, year-long design class at a large public southern university. Six case study design teams were drawn from two iterations of the course (2006-07 and 2007-08). Design teams were composed of four students, organized by course instructors, and persisted across two consecutive semesters. Teaching assistants played a large role in facilitating learning, spending approximately 100 contact hours with teams. The course comprised a two-month redesign project followed by eight-month industry-sponsored project, during which teams were mentored by faculty advisors and sponsors. Course instructors used team rankings of sponsored projects to match teams to sponsors; when multiple teams ranked a project highly, the sponsors were asked to "hire" based on resumes submitted by the students. The projects came from hospitals, industry, government, and universities, and while they varied in terms of difficulty, all were real-world, complex, and ill-structured.

## Methodology

Guided initially by theories (Fetterman, 1989) related to design process (Jonassen, 2000; Schön, 1983) and learning (Hatano & Inagaki, 1986), I entered the design studio as a non-participant observer (LeCompte & Schensul, 1999) intending to study design *learning* and to model how innovative design products might arise (Svihla, 2009). As I built rapport and reconciled my etic understandings of practices in relation to these theories, I transitioned to a participant observer role (e.g., discussions of statistical analysis, assisting with presentation preparation decisions), and began co-constructing emic understandings related to the practices and experiences of specific student design teams. Adopting an ethnographic approach (LeVine, 2007) to understanding students' experiences learning to design afforded an opportunity to consider concurrent shifts in identification.

Given the diversity of projects, comparison was of primary importance, allowing me to select cases by comparison, spontaneously use comparisons as a means to understand, and comparing across cases (Vogt, 2002). For each class iteration, I collected qualitative data related to three case study teams, providing six case study teams. Teams were selected with input from the professors and teaching assistants, who listed their highest and lowest performing teams, which teams sought out resources and which did not, and were invited to suggest teams they thought might be interesting to study for other reasons. These data along with peer evaluations and a survey of interactions formed a matrix yielding three levels (high, medium, and low performing), from which teams with all or most members consenting into the study were selected.

Teams were observed as they met together and with mentors. Meetings were audio rather than video recorded due to limitations posed by the settings: much of the design activity took place in a room with only one available electrical outlet, and this was commonly in use by a student; additionally, not all sponsors were comfortable with video recording of materials covered by Intellectual Property. Informal interviews/discussions emerged during observation, either instigated by myself or by the students. Such conversations were used for member checking and clarification, though the latter was rarely needed as the students tended to assume that I had insufficient understanding of their projects, and would often teach me and ask if I had questions.

Field notes and transcriptions of interactions were reviewed multiple times and initially analyzed as team narratives of designerly solutions (Cross, 2001) to impasses in design process (Svihla, 2009; Svihla, Petrosino, & Diller, 2009). Initial analysis of transcripts involved grounded coding, allowing themes related to initial research goals to emerge, then seeking counterexamples. This process uncovered other persistent themes related to leadership and underrepresentation in relation to the adoption of designerly perspectives. Transcripts were re-analyzed using the framework presented by Stevens and colleagues (2008), framing students' experiences by accountable disciplinary knowledge (ADK), navigation, and identification.

## Navigation

Though the focus of this study is on how students identify during their capstone design experience, it is useful to consider both unofficial and programmatic aspects of navigation preceding the capstone course. The majority of students are admitted to the degree as freshmen when they are admitted to the university, and therefore recognized as "BMEs" from their first day at the university. The program is competitive even within engineering, with the 100 highest ranking students admitted each fall. This ranking is articulated as follows: "To be competitive for admission to the Biomedical Engineering program a student needs to be very strong in the following areas: 1) Class Rank and 2) SAT composite score. Incoming First Year students are ranked based on their scores in these areas and then the students at the top of this list are offered admittance to the Biomedical Engineering degree program." Reflecting on this rather general process that selects them for membership as BMEs, one student expresses a matched generality: "When I joined BME I just picked it out of a list. It wasn't until three years into the program that I realized it was exactly what I wanted to do with my life."

After two years of core curriculum, students select a track: Biomedical Imaging & Instrumentation, Cell & Bio-molecular Engineering, Computational Biomedical Engineering, and Pre-medical Engineering. Most students opt for Cell & Bio-molecular Engineering track. These official pathways mask the unofficial navigation students undertake, which comprise myriad experiences: internships of varying relevance, engineering clubs and camps in high school, and participation in social (formal and informal) aspects of being a BME.

## Identity, Identification

Wortham (2004) describes the process of social identification as possessing characteristics or behaviors that allow others to classify an individual according to a socially agreed-upon category. This process is interactional (Gee, Staff, & Paul, 1999) and contextualized by metadiscourses (in this case, biomedical engineering students and their various mentors), but also backgrounded by other social identifications that become relevant. I employ an understanding of identity that is social, negotiated, and *double-sided*, meaning that an individual may position herself or himself in a certain manner, but this occurs within social contexts such that the individual may or may not be recognized or be positioned by institutions, organizations, or others (Holland, Lachicotte, Skinner, & Cain, 1998; Skinner, Valsiner, & Holland, 2001). Considering practices of identification (Becker & Carper, 1956) affords a focus on the actions and behaviors related to positioning. I adopt this double-sided understanding of identification to explore how students negotiate and coordinate multiple identities as they transition from problem-set engineering to engineering design.

## Accountable Disciplinary Knowledge

In their framework for understanding pathways through engineering education, Stevens and colleagues (2008) refer to *accountable disciplinary knowledge* (ADK) as one of three dimensions. This term encompasses that which is counted as engineering knowledge relative to the context. As the context shifts, the ADK shifts. Reflecting on the core coursework, one student describes a burgeoning awareness both of his chosen discipline and a sense that some of the ADK of the core coursework was of less relevance for his future:

> "What do you need to know that's actually going to be useful in the future? Because, I mean, some classes, you kind of – especially when in the first couple of semesters for BME, we kind of got to see what we were going to be doing in the future. See, I mean, and I didn't do it as much at first, but then I kind of realized, like, you know, there's certain classes that we have to take as part of the core curriculum that really aren't – parts of it I don't think matter as much. So some stuff you just learn for the test and then there's some stuff you learn for life."

In core engineering coursework, ADK includes problem sets, whereas in the design courses, which reflect the professional engineering design studio, this includes actions that reside within the educational context despite originating from the professional context, as well as actions recognizable as professional. This semi-professional setting is captured is a student's description of the design class as "working more in the 'real world' environment, and getting the chance to apply what we've been studying in the classroom." The design projects were punctuated by tools (e.g., Gantt charts, Pugh charts) borrowed from the design studio but used such that the timing may not have highlighted their relevance, as evidenced by statements that the tools "detracted from the progress of the project," and "have actually hindered us." Other tasks, such as meetings with industry sponsors and progress reports for sponsors greatly reflected the professional design studio, requiring students to negotiate professional relationships, and sometimes choose between a professional goal and a classroom goal. Such professional connections, which sometimes deviate from the course structure, can be motivating (Magleby, Sorensen, & Todd, 1991), as expressed by a student: "We were very motivated in this project by our sponsors, who both impressed on us the potential for our device and made many valuable contributions to its design." These relationships may offer points of entre to engineering or provide students with examples of professional practice, allowing them to focus on what lies beyond their coursework (Stevens, et al., 2008).

Lectures given during the capstone design course touched on the nature of BME: In one lecture the course professor, Dr. Davies (a pseudonym) described his role in helping to define this relatively young discipline ("BME is a new breed"), and encouraged the students to "be a part of the definition." These lectures served as touchstones for students to decide whether they fit with this definition, ("You are probably motivated by the practical") and to highlight the shift in ADK that accompanies the transition to capstone design:

- "Engineering is doing"
- "Solving new problems in new ways"
- "Engineering is opportunity to be creative and come up with new ideas"

These lectures included discussions about differences between science and engineering, and positioned engineers as "doers." Students were told that "engineering is design" and that as mathematics is used as a tool by scientists, science is a tool of engineering. The nature of science was represented as unproblematic (e.g., "In science, there is always a right answer") and this was reflected in the engineering science courses completed previously, filled with problem sets with right answers. In an engineering program in which approximately 30% of the students plan to go on to medical school, this divide between ADK for college and for their professional futures may lie along multiple dimensions.

## Leadership and Underrepresentation

Within the design teams, students negotiated their roles. Some teams appointed leadership roles, some leaders delegated further roles, and some teams allowed leadership to be an emergent property. In some teams, emergent leadership subsumed prior leadership roles: in each case, this occurred in conjunction with shifts in relevant expertise: leaders appointed during the preliminary project initially maintained their roles in the sponsored project, but as the ADK within the team changed, leadership shifted. In one team, this shift was explicitly recognized by the team as a whole; in the other two teams, no one claimed new leadership roles, yet in both specific individuals came to be recognized within the team as the person to check with on major decisions, though in one case (expanded upon in the second case study below, team 3.4) this role was contested.

In all case study teams, underrepresented groups, specifically women and Latinos, served in non-leadership roles. In several case study teams students from underrepresented groups were the first to promote designerly ADK. Two case studies highlight how leadership and identification with engineering or health sciences promoted types of ADK and intensified or mitigated a sense of underrepresentation. Data presented here are brief due to page limitations, however, analysis of the fuller corpus of data supports these findings.

## Team 3.2: Shifts towards Designerly Accountable Disciplinary Knowledge

In team 3.2, Tom, Greg and Cynthia all planned to become biomedical engineers at the beginning of the design class, whereas Addai explained his interest as "consulting or energy policy … Of course I might wind up working in BME, although I would be a little surprised at that." The design project involved creating a way to measure movements in a patient's limb, and was sponsored by a physical therapist who described himself as "naïve" to the nuances of the project and "not an engineer!"

Tom led his teammates by deliberately not "micromanaging" so as not to risk ending "up doing almost nothing as far as my real contributions to the engineering." Tom identified strongly with engineering, and expressed this as part of his past, both in high school and in an undergraduate internship, as well as in his future plans to get a PhD in biomedical engineering:

> "I feel the senior design is sort of just one in a series of such projects. So like in high school, I did these robotics competitions that every year […] My friend and I started this club, so we tried to recruit as many of the nerdy kids [as] we could, and there was like 30 of us, 50 of us, at various levels of involvement, different schools."

Tom's identification with engineering, beyond "nerdy" was tied to a deeper understanding of design process than many of the students expressed: **"**There's all these different solution paths, right? There's all these different ways that I can come to a solution. And then there's not even one ending. There's a lot of different places to go and then even more ways to get there. And so you're trying to choose a path." Although Tom embraced the complexity of the design space, he struggled to lead his team towards choosing "a path" because of a theoretical/engineering science framing initially adopted. Tom invited his team mates into this problem space by offering an exemplar that was accessible, as evidenced by many conversations following it, an example of which is presented below:

> Cynthia: I seriously just learned more than I did in my *entire semester* of physics.
> Tom: Very difficult.
> Addai: It's such a weird notion that you can be moving-
> Cynthia: *Yeah.*
> Addai: -and have the same vector sum as *not* moving.
> Tom: Right.
> Cynthia: That is a weird thing.
> Tom: That's, that's the whole problem with these accelerometers.

Overcoming this "problem with these accelerometers" required creativity, as Tom explained and "one of the things that spawns creativity and innovation is having so many different perspectives work at once." Not every team member can productively be a leader, and it is important to explore the experiences of the less vocal team members. Cynthia, who rarely spoke during team meetings, explained mid-way through the project that she felt she had less to contribute than her team mates. Cynthia contributions, which helped to steer the team in designerly directions by invoking the customer needs and practical requirements of the sponsor, did not reflect the ADK associated with their prior coursework, but her team mates recognized that her contributions did reflect the ADK of the design course. Tom, in particular, valued Cynthia's contributions, though this was not apparent to Cynthia until later in the design project, when the shifts in team ADK meant that her design perspectives were of particular relevance. These shifts were brought on, to some extent, by the need to generate a prototype that would help them convey to their sponsor some of the theoretically challenging aspects of the design goals (Svihla, et al., 2009). Their sponsor expressed satisfaction with their final project: "I really see even more clearly that it's a viable possibility to create this thing and before, it was, before, I had a dream list of three things or four things that I would like to see become three dimensional and I think this is phenomenal! Because it really is- they're bringing a lot more data to this than has ever even been questioned." All members of team

3.2 expressed satisfaction with the design experience, including Cynthia ("I had a great experience!") and decided to pursue graduate degrees in engineering, even Addai, who explained "It is a shame that this class is the first course I've ever taken in my almost 5 years here where I have thought I know what a biomedical engineer is, and I wouldn't mind being one."

## Team 3.4: Tensions between Pathways

Team 3.4 was led by Steve, who completed an internship working with medical testing studies on animals, a connection he leveraged to position the project as more of a medical project. Steve and Dillon both planned to attend medical school, whereas Daniela and Bob planned to become engineers. This divide was a significant one for the team, particularly given that what their sponsor desired was for the team to demonstrate, through animal testing, the feasibility of placing an existing externally-used sensor internally to monitor chemical changes related to a medical condition. During their weekly team meetings with their teaching assistant, Michelle, it became increasingly apparent that the team was not designing anything. Michelle urged them to adopt a design/engineering perspective, and tensions emerged across these disciplinary identifications. Several interactions highlight these tensions, as Daniela struggled to interject a design perspective. In the first example, Daniela posed the question to Michelle, turning to face her and speaking to her as she interjected this dissonant perspective:

Daniela: I just, um, something bothers me. The fact that we are putting the sensor on the stomach during surgery, but then we're gonna, the surgery only lasts one to two hours and we're gonna take it off and the patient is gonna be, well, the surgery is gonna be over and there's not gonna be any monitoring afterwards, and I'm thinking well, there's higher chance of sepsis, I mean shock, afterwards, right? So, should we think about leaving the sensor, or… cause I don't really?
Dillon: Seriously, that could be, like, the next project.
Steve: Yeah, I think that, like are you talking about in real life? Like…?
Daniela: Yeah, like, so using it.

Steve's question about how the device would be used ("Are you talking about in real life?) not only challenged the design perspective Daniela attempted to adopt, but also demonstrated a disconnect in perspective; Steve framed this as a school problem even though they were asked to solve a real-world problem. The response Daniela received from her team mates seemed like an attack, but Daniela did not appear to take it personally, though whenever she posed these dissenting design ideas, she tended turn her body towards Michelle and to speak facing Michelle, as if this was how she gained voice. As they continued this conversation, it became increasingly clear that there was a lack of agreement about what they were doing. They retreated into an explanation of what they were "supposed" to be doing, also highlighting a lack of design perspective:

Bob: I thought the project was to do an internal sensor that you left in?
Daniela: So we are gonna?
Steve: I think that that's, be-
Daniela: How long are we gonna leave…?
Bob: I'm not sure, uh.
Steve: I think that would be something left up to surgeon or something, honestly, likelikelike, our project, I think it's kinda outside the scope of our project.
Bob: If we left it up to the surgeon and whoever actually designs the sensor?
Steve: Yeah, whoever is *really* doing this.
Bob: 'cause we're not supposed to be designing anything.
Steve: Yeah, we're just seeing if you can do it. We just have two types of sensors and we're gonna see if we can do it- we're gonna see if a [specific type of] patient- whether or not the [chemical] levels can be measured or change to a degree that we, they show up, or the…
Bob: Using currently available sensors.
Daniela: I don't even know if it's okay to just leave it there.
Steve: It's all right.
Daniela: These are, I mean, sort of? *Days?*
Dillon: They're not gonna want to cut them open again and just take it out.

Bob and Daniela's comments express, to varying degrees, dissatisfaction with the direction their project was headed, but these practical, engineering perspectives were quashed by Dillon and Steve who together, positioned these questions as peripheral or irrelevant. Later, Bob and Daniela discussed their project and the lack of design or even engineering focus:

Bob: Well, this project is definitely nothing like the hard-core engineering I thought engineering would be like, where you go, and designing a machine, to go through those, or doing material analysis? But this is a lot more like a bioinformatics problem with the evaluation method, doing statistical analysis of various sensors, correlating how they work.
Daniela: It's not really- we're not designing it, I mean, *anything*. It would be more design if we had to design the sensor, which I thought we were gonna have to do, but he didn't want, the company doesn't want us to.

Michelle leveraged this dissatisfaction to challenge Steve again to adopt an engineering perspective:

Michelle: You have to remember how you talked about all the technical aspects. I mean, most groups are having that problem anyway, and I think, like, that would be a good project for this class.
Steve: A grade.
Michelle: I mean, if that's not what he wants-
Bob: The sponsor said 'no.' He doesn't even want us to try.

This exchange highlights tensions between the ADK of the design course/perspective, and the ADK in the project Steve thought would satisfy the sponsor. Steve identified with his prior animal studies experiences, denigrating the designerly perspective as "a grade." He enthusiastically valued the opportunity to position the design project as a continuance of his prior experiences: "The whole senior design course was worth it to me because we went through this process where something, I'm gonna be doing time and time and time again." This is starkly contrasted with his frustration with the designerly ADK:

Steve: It's just my biggest, my biggest frustration with this is just, like this, like this is the project, like *this is the project we were given*. I mean if, if like Dr. Davies, and I don't mean to be disrespectful, I mean, if somebody didn't want us to do this project it shouldn't have been accepted. It shouldn't have been given to us.

Steve constructed a science problem space and phrases such as "project we were given" indicate that a lack of flexibility with regard to that problem space. That statement in particular is troubling because it is the designers' job to define that problem space. Whether he recognized it or not, they did, in fact, create a problem space, but one that was framed as a science rather than engineering problem.

Steve brought the focus back to the novelty of their project's science goals, at least in terms of the course. The novelty that excited Steve, troubled Michelle and Dr. Davies because it was effectively a substitute for a design project. As Steve expressed this, Daniela once again turned to Michelle and expressed a dissenting designerly perspective:

Steve: I mean, like a surgical procedure? I mean, that's not considered any sort of technical…? Like, a surgical procedure? And, like, taking readings in a rat's stomach? I mean, something that nobody else has *even* ventured to do in *this* course?!
Daniela: Well, we're not *designing* anything.

Because of concerns Dr. Davies met with the team. He explained to them:

> "Now, in BME, we have to be careful a little bit. It's really easy to get off and to, uh, do life sciences kind of things that are not engineering, and nearly any project can, uh, have identified some appropriate engineering dimensions for it and, uh, I think this is an important part of your education experience. […] I encounter this in my own research and can identify all sorts of neat medical things to work on. The question is, *What can you really uniquely contribute as an engineer?* Somebody across the street, this brilliant life scientist, is not gonna be able to put on the table?"

Though the sponsor was very impressed with what the team accomplished, not all team mates felt enthusiastic about the experience: "There was no design portion to the sponsored project. The design portion was assigned by Dr. Davies halfway through the project," and furthermore, this aspect was "created to fulfill graduation requirements."

## Discussion

Viewing these cases of design team lived narratives through the dimensions of accountable disciplinary knowledge, navigation, and identification provides ways to understand how differing pathways create tensions in capstone design, but not earlier. When the ADK is no longer the domain of problem sets, but instead reflects the professional design studio, students who have no intention of becoming engineers have opportunity to exploit the lack of designerly perspective brought on by having an official pathway containing no previous design experiences (Figure 1). Given that many students, as part of their unofficial navigation, have worked in laboratories doing what Dr. Davies termed "life sciences kind of things that are not engineering" it is unsurprising that a team leader might consider such activities to be the ADK of the design studio, and to see them as valuable in the future. This is expressed by Steve, reflecting on the animal testing his team undertook: It "was worth it to me because we went through this process where something, I'm gonna be doing time and time and time again." Steve expected to find the ADK of medical school to have more in common with his prior coursework ADK than that of the design course, and this made the design aspects very artificial ("A grade.").

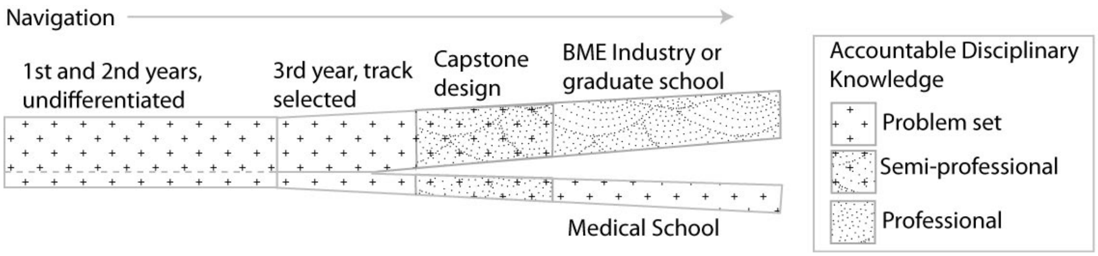



Figure 1. Changes in accountable disciplinary knowledge along pathways towards engineering and medicine.

In contrast, Tom, whose unofficial navigation included very different experiences, desired to adopt designerly ADK, though he struggled to do so. That it was such a challenge even given his desire and his team's desire to engage with design highlights what an opportunity Steve had to steer his team towards a "life sciences kind of thing." Given that official navigation pathways lack early design experiences, it is left to those with design experience from unofficial navigation to be the *chauffeurs* of designerly perspectives. In these case study teams, this role was filled by women. Thus gender emerged as a complex dimension because these women who helped to shift their teams from problem-set ADK to designerly ADK, viewed these designerly perspectives as somehow weaker than problem-set ADK. This perspective is warranted by their prior experiences in courses that emphasize engineering science. Though women were not the *chauffeurs* of designerly ADK in all case studies, gender is an important aspect to consider, given that problem-set ADK holds a privileged position garnered through prior experiences; this, taken with the degree to which team mates, especially those in leadership roles recognize the need to make such a shift, may explain, in part, why Daniela's repeated appeals for designerly perspectives were left largely unheard, whereas Cynthia's were adopted.

The official navigation creates a space in which those who champion designerly ADK may be positioned within a design team as having productive, disruptive, or irrelevant perspectives. When the chauffeurs of designerly ADK are from groups underrepresented in engineering, there is the potential to widen or narrow gaps, depending on whether those perspectives are treated as productive, disruptive, or irrelevant.

Institutional identification of students as BMEs from the first day has been noted as beneficial elsewhere (Stevens, et al., 2008). Taking identification as double-sided, it is relevant to consider students may identify themselves as engineers or as pre-medical, and that for those who identify themselves as engineers prior to the capstone design, official navigation offers few opportunities to reconcile what this means. Adoption of designerly ADK depends much, therefore, on endorsement by team leaders and unofficial navigation.

## Implications

The focus on engineering science in prior coursework and accompanying accountable disciplinary knowledge places much of the burden of becoming an engineer upon the capstone design course. The open-ended nature of the design projects, the lack of prior design experience, and the legitimacy of BME as pathways to both engineering and medical school means that students may steer their projects towards or away from engineering design projects. If students so not see design as relevant to the pathway they have selected, they are unlikely to engage with a project as designers, and may even discourage their team mates from adopting designerly perspectives. For these students, identification as engineers is contingent indeed.

## References

Becker, H., & Carper, J. (1956). The development of identification with an occupation. *The American Journal of Sociology, 61*(4), 289-298.

Bucciarelli, L. L. (1994). *Designing Engineers.* Cambridge, Mass: MIT Press.

Cross, N. (2001). Designerly Ways of Knowing: Design Discipline Versus Design Science. *Design Issues, 17*(3), 49-55.

Fetterman, D. M. (1989). *Ethnography: Step by step.* Newbury Park: CA: Sage Publications.

Garrison, L., Stevens, R., & Jocuns, A. (2008). *Gender, institutional structure and learning in an engineering college*.
Gee, J., Staff, G., & Paul, J. (1999). *An introduction to discourse analysis: Theory and method*: Routledge.
Goldschmidt, G. (1996). The designer as a team of one In N. Cross, H. H. C. M. Christiaans & K. Dorst (Eds.), *Analysing design activity* (pp. 65-91). Chichester, UK: Wiley.
Gunther, J., Frankenberger, E., & Auer, P. (1996). Investigstion of Individual and Team Design Processes. In N. Cross, H. H. C. M. Christiaans & K. Dorst (Eds.), *Analysing design activity*. Chichester: Wiley.
Guzzo, R. (1986). Group decision making and group effectiveness in organizations. In P. S. Goodman (Ed.), *Designing effective work groups* (pp. 34-71). San Francisco, CA: Jossey-Bass.
Hagström, T., & Kjellberg, A. (2000). Work values and early work socialization among nurses and engineers. *Health effects of the New Labour Market. New York: Kluwer Academic*.
Hatano, G., & Inagaki, K. (1986). Two courses of expertise. *Child development and education in Japan*, 262–272.
Heyman, G., Martyna, B., & Bhatia, S. (2002). Gender and achievement-related beliefs among engineering students. *Journal of Women and Minorities in Science and Engineering, 8*(1), 41.
Holland, D., Lachicotte, W., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard Univ Press.
Jonassen, D. H. (2000). Toward a Design Theory of Problem Solving. *Educational Technology Research and Development, 48*(4), 63-85.
Kidder, D. L. (2002). The Influence of Gender on the Performance of Organizational Citizenship Behaviors. *Journal of Management, 28*(5), 629-648.
LeCompte, M. D., & Schensul, J. J. (1999). *Designing & Conducting Ethnographic Research* (Vol. 1). Walnut Creek: AltaMira Press.
LeVine, R. (2007). *Culture, behavior, and personality: an introduction to the comparative study of psychosocial adaptation*. Chicago, IL: Aldine De Gruyter.
Magleby, S., Sorensen, C., & Todd, R. (1991, November). *Integrated product and process design: a capstone course inmechanical and manufacturing engineering*. Paper presented at the Frontiers in Education Conference.
Ochs, E., & Capps, L. (2001). *Living narrative: Creating lives in everyday storytelling*: Harvard Univ Pr.
Phelps, E., & Damon, W. (1989). Problem Solving with Equals: Peer Collaboration as a Context for Learning Mathematics and Spatial Concepts. *Journal of Educational Psychology, 81*(4), 639-646.
Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
Schreuders, P., Rutherford, B., Cox, K., & Mannon, S. (2006). *Gender differences in biological engineering students*.
Seymour, E., & Hewitt, N. (2000). *Talking about leaving: Why undergraduates leave the sciences*: Westview Press.
Skinner, D., Valsiner, J., & Holland, D. (2001). Discerning the dialogical self: A theoretical and methodological examination of a Nepali adolescent's narrative. *Forum: Qualitative Social Research, 2*(3). Retrieved from http://nbn-resolving.de/urn:nbn:de:0114-fqs0103187
Stevens, R., O'Connor, K., Garrison, L., Jocuns, A., & Amos, D. (2008). Becoming an Engineer: Toward a Three Dimensional View of Engineering Learning. *Journal of Engineering Education*.
Svihla, V. (2009). *How differences in interaction affect learning and development of design expertise in the context of biomedical engineering design*. The University of Texas, Austin.
Svihla, V., Petrosino, A. J., & Diller, K. R. (2009). *Learning to design: Interactions and collaboration* Paper presented at the AERA, San Diego.
Vogt, F. (2002). No ethnography without comparison: the methodological significance of comparison in ethnographic research. In G. Walford (Ed.), *Debates and developments in ethnographic methodology* (Vol. 6, pp. 23-42). Oxford: Elsevier.
Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87*(3), 406-423.
Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*: Cambridge Univ Pr.
Wortham, S. (2004). The interdependence of social identification and learning. *American Educational Research Journal, 41*(3), 715.

## Acknowledgments

This paper was greatly improved by comments from reviewers and Reed Stevens.

# An Overview of CSCL Methodologies

Heisawn Jeong, Hallym University, Chuncheon, Korea (South), heis@hallym.ac.kr
Cindy E. Hmelo-Silver, Rutgers University, New Brunswick, US, cindy.hmelo-silver@gse.rutgers.edu

**Abstract:** This study surveyed empirical CSCL papers published in seven leading journals of the field during 2005-2007 and analyzed methodologies of a random sample of 33 studies. The dominant features of CSCL research methodologies were descriptive studies in classroom settings with self-report/questionnaire that were quantitatively analyzed. Yet, studies were quite eclectic in their research method, adopting a variety of research design, data sources and analysis methods, leading to cross-over or hybrid methodologies. Studies of collaborative learning were carried out frequently both in experimental and descriptive studies in lab and classroom settings. Diverse data sources were collected and analyzed both quantitatively and qualitatively. New techniques for analyzing collaborative processes were also emerging. Still, it appears that the field needs to be more principled in its applications of design-based research. In addition, the field is in need of more diverse and systematic ways to analyze qualitative data in order to take a full advantage of the rich data afforded in CSCL research.

## Introduction

There has been an increasing interest in collaborative knowledge construction for some time. With the advent of information technology, collaboration is also increasingly mediated by various technologies. Computer-Supported Collaborative Learning (CSCL) is an interdisciplinary research field that are concerned with how people can learn together with the help of computers (Myake, 2006; Stahl, Koschmann, & Suthers, 2006). In any field including CSCL, methodology is of utmost importance in advancing the field. Methodologies are constrained by the current level of understanding and available analysis tools in the field, and yet advancements in methodology can also open up new conceptual spaces and allow researchers to examine phenomena and questions that were not previously possible. Reflecting its interdisciplinary mission, diverse methodologies co-exist in CSCL. Traditional methods initially developed in the context of studying individuals (e.g., questionnaires, study of individual problem solving protocols) have been actively used by CSCL researchers. CSCL researchers also have reached out to methodologies in fields such as linguistics and anthropologies. The infusion of these methodologies has helped to advance the field a great deal. Currently, however, it is difficult to understand what the current state of research methodologies is in CSCL. Numerous methodologies co-exist without a clear understanding of what they are and how they are related to each other.

Recently, several meta-analyses have been conducted in related fields of CSCL. Hrastinski and Keller (2007) examined research approaches (e.g., empirical versus conceptual studies, quantitative or qualitative methods, etc.) of papers published in four leading journals of *educational technology* during 2000-2004 period. They found that about two thirds of the studies were empirical investigations, about half of which (51%) used quantitative methods, 25% used qualitative methods, and 24% used mixed methodologies (Hrastinski & Keller, 2007). Hew, Kale, and Kim (2007) also conducted similar meta-analysis on research topics and methodologies in the field of *instructional technology* based on publications in three journals also during 2000-2004. They focused on empirical articles and found that descriptive and/or correlational studies were the dominant research methods in educational technology, being used in more than half of the studies published. Methodologies varied somewhat depending on research topics; experimental method was more commonly used in studying the topic of psychology of learning and teaching, while correlational method was most frequently used with the topic of research and evaluation methodology. In addition, the studies most commonly collected survey/questionnaire data in higher education and K-12 settings (Hew et al., 2007).

Distinctions such as quantitative and qualitative methodologies have long existed in many fields. Quantitative methods were typically associated with experimental and survey studies where numerical data are collected. Analysis aimed to uncover 'general' trends that are true to the population at large, which lead to the focus on large sample sizes and quantitative analyses using inferential statistics. Qualitative methods, on the other hand, were associated with descriptive research such as case or ethnographic studies where qualitative data such as video-tapes, verbal transcripts, and artifacts are collected. Analysis aimed at uncovering 'meaningful' patterns and in-depth look at a small set of data was emphasized. While specific methodological traditions in qualitative research may focus on different aspects of the data and may take different routes to arrive at their interpretations, data were mostly analyzed qualitatively. Such a distinction, however, no longer seems appropriate. As qualitative methods become more popular, more attempts are made to triangulate data collection and analyses. Traditional descriptions of qualitative or quantitative research no longer adequately describe some of the research being carried out in the field. There is a need to develop a more sophisticated understanding of what quantitative and qualitative methods mean and how they can be better integrated to inform CSCL research at large. In addition, we need to examine and evaluate how the current CSCL methodology is addressing some of the unique challenges that CSCL faces. With its emphasis on social construction of knowledge mediated through technological tools, CSCL has been in need of methodologies to study learning as it happens in the process of interacting with learning partners and tools. Interaction is a complex phenomenon where traditional methodologies developed to study individual learning are not necessarily readily applicable. Moreover, interaction in CSCL environments is unique in that there exist diverse forms of interaction such as distributed as well as face-to-face interaction and that interaction is often mediated by technological tools (Lehtinen, Hakkarainen, Lipponen, Rahikainen, & Muukkonen, 1999). Because thinking often develops and is made observable through dialogues and discourse, research efforts have been directed to the analyses of dialogue and discourse both in face-to-face distributed interaction. Analysis methods such as content analysis, discourse analysis, and conversation analysis have thus become commonplace in CSCL literature. Combining methodologies from different research traditions has been useful for revealing ambiguities and contradictions, which in turn have led to new conceptual developments. However, it have increased the risk of confusion in different epistemological perspectives (Naidu & Jarvela, 2006). Individual researchers are confused about how different methodologies can be used to inform their research. It is equally unclear whether CSCL methodologies have been adequately addressing the key issues of CSCL.

In response to these concerns, this study attempted to provide an overview of current methodologies used in empirical investigations of CSCL. We first carried out a comprehensive examination of empirical CSCL studies, analyzing research methodology at the following four dimensions: (1) research designs (e.g., descriptive or experimental design), (2) study settings (e.g., classroom or laboratory), (3) data collected (e.g., survey data, log data, or synchronous or asynchronous text data), and (4) analyses carried out with the data (e.g., content analysis, social network analysis, multi-level analyses). Based on this analysis, we also examined how collaboration is studied in CSCL research.

## Methods

### Journal Selection

*International Journal of Computer Supported Collaborative Learning* (ijCSCL) is the flagship journal of the CSCL community. In addition to ijCSCL, there are numerous other research journals where CSCL research is published. We asked the leaders of CSCL community, CSCL committee of ICLS and the editorial board members of ijCSCL to nominate five leading journals of the field other than ijCSCL. Based on the responses from 16 community leaders, the following seven journals were selected: (1) *ijCSCL*, (2) *Journal of the Learning Sciences,* (3) *Learning and Instruction*, (4) *Computers and Education*, (5) *Journal of Computer Assisted Learning*, (6) *International Journal of Artificial Intelligence in Education*, and (7) *Computers in Human Behavior.* They are all peer-reviewed journals published by well-known publishers with international authorship and readership.

### Paper Selection

From the papers published in the seven journals, we identified *empirical CSCL* papers to be used for the content meta-analysis. By *empirical*, we mean that the study collected and analyzed primary data. Secondary data analysis, simulated results, theoretical papers and meta-analyses were not included. The data may have been collected as part of a larger project, but the analysis and finding had to be new. Papers that described the design process were included if they reported on empirical data. By *CSCL* research, we mean studies where students learned collaboratively using computers or other technological tools. Learning needed to be collaborative, but it does not mean that all phases of learning needed to be collaborative. As long as parts of the learning process involve interaction (e.g., collaborative discussion after individual study), it was considered collaborative. We focused on small group peer collaboration, that is, collaboration among learners who are similar in knowledge and status. This means that studies that examined student-teacher interaction or whole class discussions were not included unless they included small group peer collaboration. The applied technologies do not necessarily have to be collaboration technology such as e-mails or discussion boards, but need to be specific. Interaction with computerized agents were included if it involved learning. Studies about motivation or attitudes were included if they were studied in relation to learning. In addition to empirical CSCL papers, we included methodological papers that addressed various methodological issues related to CSCL (e.g., introduction of new methods such as Social Network Analysis, development of specific rating schemes).

Excluding non-research articles (e.g., editorials, book reviews, or obituaries), 868 articles published in the 2005-2007 period were screened, which means three years of publication (two years of publication in the case of *ijCSCL*). The total numbers of selected papers is currently 175, indicating that 20% of the papers published in the seven journals during 2005-2007 were empirical or methodological investigations of CSCL.

## Content Analyses

The following aspects were coded for each paper: (1) Design, (2) Study setting, (3) Data, (4) Analysis, and (5) Interaction. In this paper, we report on the analyses carried out on a random sample of 33 papers. This set constituted 21% of the total empirical CSCL paper (and 19% of the total CSCL papers including methodology papers; percentages did not come out even due to the ongoing nature of paper selection process) and were marked with * in the reference list. A combination of inductive and deductive approaches was used to develop coding categories for each code. Coding categories were initially generated top-down (e.g., using categories drawn from the submission descriptors of the 2005 CSCL conference) and then later refined bottom-up through multiple iterations of coding. Jeong and Hmelo-Silver (2010) reported on a different set of analyses (with the exception of the collaboration coding) carried out on the same sample. Specific codes for each category are described below.

### Research design

Research design refers to whether the study is (1) Experimental, (2) Descriptive, or (3) Design-based method. Experimental design refers to studies where variables are manipulated and was further divided into (a) randomized, (b) quasi-experimental, and (c) pre-post design. Descriptive studies do not manipulate variable and assign study participants to different conditions. They study the variables and phenomena as is and include studies such as case studies, surveys, and ethnographic investigations. Design-based method refers to research strategy where CSCL designs and interventions are theoretically-driven and are refined progressively over several iterations. In order to be coded as design-based method, the study not only needs to design CSCL systems or environments, but also the design needs to be theoretically grounded, instantiated in specific contexts, studied and refined iteratively as part of a bigger design-based research (Barab & Squire, 2004; Brown, 1992; Collins, Joseph, & Bielaczyc, 2004). Once a study was coded as design-based method, the design of individual iterations which can be either experimental and/or descriptive was not coded separately.

### Study settings

Study setting refers to the contexts in which the research was conducted, that is, whether the research was carried out in (1) Laboratory, (2) Classroom, or (3) Other settings (e.g., informal learning environments or workplaces). Laboratory means that learning occurred in lab or lab-like controlled settings where the task occurred isolated by itself outside the context of classroom or other authentic learning situations. Classroom setting means that learning occurred as part of classroom/curricular activities, which could involve not only physical classroom but also other settings (e.g., field trips).

### Data

Data refer to the types of data collected in the study and was coded into (1) Process, (2) Outcome, and (3) Miscellaneous data types. Process data were further divided into (a) text-asynchronous, (b) text-synchronous, (c) video, (d) audio, (e) log data, and (f) other. When papers reported on both video and audio data (e.g., Li et al., 2006), they were coded only as the video data unless they were analyzed separately. Outcome data were further divided into (a) multiple-choice questions, (b) open-ended questions, (c) expert ratings (e.g., experts' rating of students' work), (d) artifacts (e.g., contents of multi-media whiteboard), and (e) other (e.g., final course grades). Outcome data can be collected at the beginning of a study (e.g., pre-test) as well as at the end of a course. Miscellaneous data include (a) self-reports/questionnaires (e.g., demographic information, affects, perceived acceptance, etc.), (b) interview and/or focus groups, (c) field notes or observations, (d) individual difference measures (e.g., IQ, learning styles), and (3) other (e.g., course registration information).

### Analysis methods

Analysis methods refer to the kinds of analyses carried out on the data. There were two general categories: (1) Quantitative and (2) Qualitative. Quantitative analyses were further coded into the following sub-types: (a) simple descriptive or quantitative (e.g., simple frequencies and means, as well as simple quantitative analysis of qualitative data such as coding numbers of words in messages or scoring an open-ended answers), (b) code and count, (c) inferential statistics, (d) causal or multi-level modeling, (e) sequential analyses, (f) social network analysis, and (h) other miscellaneous quantitative analysis (e.g., comparison with simulated data). Code and count refers to what is often called "quantitative content analyses" (de Wever, Schellens, Valcke, & van Keer, 2006). Qualitative analysis refers to analysis that analyzes data in a qualitative manner. Qualitative analyses were further coded into: (a) (qualitative) content analysis, (b) conversational analysis, (c) grounded theory, (d) interaction analysis, (e) loosely defined.

### Collaboration

Collaboration refers to the types of student-to-student interaction and consisted of four categories: (1) Face-to-face, (2) Distributed synchronous, (3) Distributed asynchronous, and (4) Other. Face-to-face collaboration refers to co-located and synchronous interaction. Distributed collaboration refers to collaboration among distributed learners and can be either synchronous (e.g., through chat) or asynchronous (e.g., through e-mails and discussion boards). Other collaboration refers to miscellaneous interaction such as non-verbal interaction (e.g., interaction through argumentation map), interaction with computer agents, or indirect interaction (e.g., collaborative filtering or data sharing).

Methodological features of the studies were generally coded according to the authors' definitions. If authors call their design as 'experiment' then we coded them as such. In unclear cases where the authors did not explicitly state information relevant to the coding categories (e.g., no mention of whether the study was carried out in lab or in classrooms), we relied on contextual information presented in the paper. For example when a study did not specify its setting, if 'recruited' students performed a stand-alone activity with no connection to classroom activities, its setting was coded as laboratory (e.g., Mercier & Frederiksen, 2007). Similarly, when a study did not specify data and only stated that the number of words in asynchronous notes was analyzed, it was assumed that asynchronous text messages were collected as data (e.g., Hewitt & Brett, 2007). In a few controversial cases, we followed a more conventional approach so that 'near synchronous' interaction was coded as asynchronous interaction and that an 'experiment' without any mention of treatment conditions were coded as a descriptive study. Multiple coding was allowed when more than one coding categories were applicable. For example, if the study compared different versions of a course that was implemented in face-to-face, synchronous, and asynchronous learning situations (e.g., Kitsantas & Chow, 2007), all different types of collaboration were coded. Reliability check was completed in all code (except for the collaboration code, which is still in progress) by having two coders independently code the set while discussing unclear cases. Cohen's kappa values between the two coders were all above .75 in all coding categories.

## Results

### Methodological Features of CSCL Research

The analyses indicated that the most prevalent CSCL research design is descriptive design: 58% of the studies were descriptive, 33% were experimental, and 9% were design-based. Of the experimental studies, 45% used randomized experiments, 27% quasi-experimental method, and 27% single-group pre-post design.

As for the settings of CSCL research, studies were mostly carried out in the classroom: 82% of the studies were carried out in the classroom, 12% were in the laboratory, and 6% in other settings (e.g., workplaces and informal settings). Classroom studies are typically associated with descriptive design, whereas laboratory studies are typically associated with experimental design. While this indeed was a case with most of the descriptive studies being carried out either in classrooms or other settings, there were a sizable number of studies that 'crossed over'. The majority (70%) of the classroom studies was either descriptive or design-based studies, 30% were experimental studies. Classroom experiments were typically quasi-experiments using intact classrooms or single group pre- and post-test comparisons (e.g., Cho, Gay, Davidson, & Ingraffea, 2007), but there were a few cases of randomized classroom experiments as well. Random assignments in classroom setting were carried out by assigning different treatment groups to separate laboratory sessions (e.g., Vizcaino, 2005) or by online groupings (e.g., Cho & Schunn, 2007).

As for data collection, 73% of the CSCL studies collected process data, 61% of the studies collected outcome data, and 70% collected other data. The majority of the studies collected more than one data types: Only 15% of the studies collected a single data type, and the rest of the studies collected multiple types of data. Studies on average collected 3.0 data in a given study with some studies collecting up to six different data. Of the process data, log data were the most frequently collected (42%), asynchronous text-messages were the next (30%), followed by synchronous text (15%), video-(12%) and audio-data (9%). Of the outcome data, and other types was most popular (24%), followed by artifacts (21%), open-ended (18%), multiple-choice questions (12%) and open-ended questions (12%), and experts' rating (9%). Other outcome types include data such as final grades, students' self-ratings and evaluations. Of the miscellaneous data types, self-report/questionnaire was most popular (61%), followed by interviews (21%), field notes (12%), individual difference (6%), and other (3%). A quite diverse set of data is being collected in CSCL research. Additional sources of data can provide more information about the questions. It appears that CSCL researchers are actively attempting to examine learning from multiple perspectives.

As for the analysis method, some form of quantitative analyses was carried in all studies. The most common forms were simple descriptive and inferential statistics (58% each), followed by code and count (52%); Causal or multi-level modeling was used in 6% of the articles, and social network analysis and other analysis were used in 3% of the studies each. A notable trend was a development of new quantitative methodologies such as social network analyses (Cho et al., 2007) and multi-level analyses (de Wever et al., 2006) that aimed to analyze collaborative processes quantitatively. Qualitative analyses were also used quite frequently, being used in 42% of the studies. However, the methods in general were not well described or attributed to one of the more specific genre of qualitative research such as grounded theory (Glaser & Straus, 1967) or conversational analysis (Goodwin & Heritage, 1990). The majority of the studies (93%) adopted the "loosely defined" method in which data was qualitatively described with some examples.

## Studying Collaboration in CSCL Research

How is collaboration studied in CSCL research? In terms of the types of collaboration investigated in CSCL research, 36% of the CSCL studies examined face-to-face collaboration, 36% synchronous, 48% asynchronous, and 12% other types of collaboration. The majority of the studies (70%) implemented and studied only one type of collaboration, but there were studies that implemented either two (24%) or three (6%) different types of collaboration. It was due to the fact that CSCL environments often allow students to interact with each other not only asynchronously through e-mails, but also synchronously through chats (e.g., Yang & Liu, 2007) and also because studies compared different types of collaboration in a given study (e.g., Kitsantas & Chow, 2007).

How were these various forms of collaboration studied in CSCL research? Comparisons of the studies that examined different collaboration types (see Table 1) indicated that while face-to-face and synchronous collaboration are studied equally likely with experimental and descriptive design, asynchronous collaboration were mostly studied with descriptive design. Distributed collaborations, both synchronous and asynchronous, were studied exclusively in classroom settings, less likely to collect outcome data and carry out qualitative analyses. Other types of collaboration were more likely to be studied in experiments in laboratory setting with quantitative analysis. Further analyses on the sub-type of data showed that, not surprisingly, face-to-face collaboration studies collected either video or audio data, whereas synchronous and asynchronous collaboration studies collected synchronous and asynchronous text data, respectively. Studies of other collaboration types mainly collected log data, but also collected video or other types of process data along with the log data. While quantitative analyses were used across all collaboration types, qualitative analyses were more frequently used to study face-to-face collaboration than distributed collaboration. A number of factors such as theoretical commitments, research questions, and available study settings influenced the choice of research methodologies. More analyses are needed, but it appears that the choice of CSCL research methodology is also dependent on the type of collaboration examined.

Table 1: Studying collaboration in CSCL research

| | Research Method | | | Settings | | | Data | | | Analyses | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Exp | Des | DB | Lab | Class | Other | Proc | Out | Misc | Quant | Qual |
| F2F | 38% | 50% | 13% | 25% | 75% | 0% | 88% | 100% | 63% | 100% | 63% |
| Synch | 40% | 40% | 20% | 0% | 100% | 0% | 80% | 40% | 80% | 100% | 40% |
| Asynch | 14% | 71% | 14% | 0% | 100% | 0% | 86% | 71% | 57% | 100% | 43% |
| Other | 100% | 0% | 0% | 67% | 33% | 0% | 67% | 67% | 33% | 100% | 0% |

Notes. *F2F* refers to face-to-face, *Synch* refers synchronous, and *asynch* refers to asynchronous collaboration. *Exp* refers to experimental, *Des* refers to descriptive, and *DB* refers to design method. *Proc* refers to process, *Out* refers to outcome data. *Quant* refers to quantitative and *Qual* refers to qualitative analyses. Note that there were studies that investigated more than one type of collaboration. In order to isolate research methodologies uniquely associated with different collaboration types, studies that investigated only one form of collaboration were included in the comparisons: There were eight studies that exclusively studied face-to-face collaboration, five such studies for synchronous collaboration, seven studies for asynchronous collaboration, and three studies for other types of collaboration.

Study of collaborative interaction in CSCL research is often accompanied by an accumulation of large amount of qualitative data such as video-data and discussion board messages. Such qualitative data were traditionally analyzed qualitatively most of the time using methods such as discourse analysis (Brown & Yule, 1983; Gee, 2005) or grounded theory (Straus & Glaser, 1967). How does recent CSCL empirical research approach the analyses of these qualitative data? With this question in mind, we examined how the six types of qualitative data coded in this study—video audio, text-synchronous, text-asynchronous, open-ended outcome, and artifacts—were analyzed in recent CSCL research (see Table 2). Additionally, we also examined how log data were analyzed in CSCL research. An ever increasing amount of log data is collected in CSCL research, and yet there is only a scant understanding as to how they can be used in the context of studying collaborative learning. The results indicated that the majority of the qualitative data was quantified. The quantification was mostly done using procedures such as code and count or scoring, but "simple descriptive" measures were also directly calculated in some cases, using quantitative indices of the qualitative data (e.g., word counts in notes). While most of the process data were quantified with code and count procedure (i.e., coding the presence of certain code and counting its frequency), open-ended outcome data was analyzed with scoring procedure (i.e., assigning a score using a criterion). In addition, qualitative data, especially video-data and synchronous and asynchronous text data were also often analyzed 'qualitatively', but specific kinds of qualitative analysis methods were seldom used. The qualitative analyses were mostly 'loosely defined' in that they provided qualitative descriptions of the data along with examples without committing to specific qualitative methods. As for the log data, not surprisingly, they were analyzed quantitatively, but the range of log data used for analysis was quite diverse. Studies examined time to post messages, length or duration of given threads, pages read, number of files in the folders, or participation rates. Lastly, the results also indicated that CSCL research may be good at collecting multiple types of data, but are not always good at analyzing them. Of the 50 pieces of data examined, one fifth (19%) was either unanalyzed or lacked sufficient detail as to how they were analyzed.

Table 2: Analyses of qualitative (& log) data in CSCL research

| | Simple descriptive | Scoring | Code & count | Qualitative analysis | Unanalyzed | Unspecified |
|---|---|---|---|---|---|---|
| Video | 0% | 0% | 75% | 75% | 25% | 0% |
| Audio | 0% | 0% | 67% | 0% | 33% | 0% |
| Text-synchronous | 0% | 0% | 60% | 40% | 20% | 20% |
| Text-asynchronous | 10% | 0% | 70% | 30% | 0% | 10% |
| Open-ended | 29% | 71% | 0% | 0% | 14% | 0% |
| Artifacts | 14% | 0% | 29% | 14% | 29% | 14% |
| Log data | 93% | 0% | 0% | 0% | 7% | 0% |

## Discussion

The analyses of this study showed that the typical methodology of CSCL research was descriptive studies carried out in classroom settings with the most frequently collected data being self-report/questionnaire data analyzed quantitatively. While these emerge as dominant features of CSCL methodologies, CSCL methodology is far from monolithic and can be characterized as multi-method. Studies often collected multitudes of data and also used both quantitative and qualitative analyses. As a result, research methodologies of a given study often do not fit the traditional quantitative and qualitative divide and/or experimental and descriptive divide. This seems to be driven both by the need to understand CSCL from multiple perspectives (Hmelo-Silver, 2003; Suthers, 2006) and also with the infusion from different methodological traditions. Learning, especially in CSCL contexts, is a complex phenomenon, understanding of which requires multiple approaches and perspectives ranging from individual learner, interactive processes, tool mediation, teacher roles, to general cultures. In order to achieve a richer and integrated understanding of learning, CSCL research is increasingly required to examine learning from multiple perspectives, using multiple data sources and analyses methods in the contexts of specific research questions and settings.

In spite of the dynamic developments in the field, it appears that there are a number of areas where more sophisticated methodological understanding and practices are needed. One is with what researchers called as design-based research. Although we did not carry out systematic analyses, it was clear in the coding process that researchers often did not distinguish design as a research goal from design as research methodology. Even in the case where design research was used to refer to the research method or strategies, its applications were often name in only. While exceptions exist, studies that reported on multiple iterations and progressive design refinements were rare. The field needs to be more principled in its applications of design-based research. In addition, there is a need for the field to develop a larger repertoire of analytical tools to deal with the large amount of qualitative data collected in CSCL research. Large amounts of online and offline messages and dialogues can be and are collected in many CSCL investigations. The advancement of the field greatly depends on how well these rich sources of data can be utilized. While existing qualitative methods have their values, they are not well suited to a large-scale analysis. In addition, epistemological commitments of some of these methodologies are often in conflict with those from more quantitative traditions emphasizing objectivity and generality of findings. Although qualitative analyses were used in close to half of the studies examined in this study, they rarely adopted a specific qualitative method and just described the data with no reference as to how the conclusions were derived. It might be the case that the researchers failed to attribute their analyses to specific methods even though their analyses belong to that category, but it might also be the case that existing methods may not adequately describe what researchers are currently doing with qualitative analysis. Qualitative analyses tend to be less explicit in its procedures and more challenging to learn and execute. Efforts are needed to systematize existing qualitative analyses procedures so that they are easy to understand and carry out and also to develop new analytical procedures to analyze them.

Given the small sample size of this study, the results and conclusions from this study should be taken as tentative at the moment. We aim to obtain more robust findings with additional analyses, but hope that an overview of CSCL research methodology even in this small scale is informative to researchers both in terms of understanding current CSCL research methodologies and guiding their future research. We also hope that the methodological issues raised in this study contribute to establishing a common conceptual framework for CSCL methodologies and integrating diverse findings from different research traditions.

## References

Barab, S. A., & Squire, K. D. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*(1), 1-14.

Berge, O., & Fjuk, A. (2006). Understanding the roles of online meetings in a net-based course. *Journal of Computer Assisted Learning, 22*, 13-23*.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of The Learning Science, 2*(2), 141-178.

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University press.

Chen, Z.-H., Chou, C.-Y., Deng, Y.-C., & Chan, T.-W. (2007). Active open learner models as animal companions: Motivating children to learn through interacting with My-Pet and Our-Pet. *International Journal of Artificial Intelligence in Education, 17*, 145-167*.

Cho, H., Gay, G., Davidson, B., & Ingraffea, A. (2007). Social networks, communication styles, and learning performance in a CSCL community. *Computers and Education, 49*, 309-329*.

Cho, K., & Schunn, C. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education, 48*, 409-426*.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15-42.

Cress, U. (2005). Ambivalent effect of member portraits in virtual groups. *Journal of Computer Assisted Learning, 21*, 281-191*.

de Wever, B., Schellens, T., Valcke, M., & van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers and Education, 46*, 6-28.

Ellis, R. A., Goodyear, P., Prosser, M., & O'Hara, A. (2006). How and what university students learn through online and face-to-face discussion: conceptions, intentions, and approaches. *Journal of Computer Assisted Learning, 22*, 244-256*.

Fuks, H., Pimentel, M., & de Lucena, C. J. P. (2006). R-U-Typing-2-Me? Evolving a chat tool to increase understanding in learning activities. *International Journal of Computer-Supported Collaborative Learning, 1*, 117-142*.

Gee, J. P. (2005). *An introduction to discourse analysis: Theory and method* (2nd ed.). New York: Routledge.

Glaser B.G. & Strauss A. (1967). Discovery of trounded theory: Strategies for qualitative research. Hawthorne, NJ: Aldine Transaction.

Goodwin, C., & Heritage, J. (1990). Conversation analysis. *Annual Review of Anthropology, 19*, 283-307.

Hakkinen, P., & Jarvela, S. (2006). Sharing and constructing perspectives in web-based conferencing. *Computers and Education, 47*, 433-447*.

Hew, K. F., Kale, U., & Kim, N. (2007). Past research in instructional technology: Results of a content analysis of empirical studies published in three prominent instructional technology journals from the year 2000 through 2004. *Journal of Educational Computing Research, 36*(3), 269-300.

Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *The journal of the Leaning Sciences, 14*(4), 567-589*.

Hewitt, J., & Brett, C. (2007). The relationship between class size and online activity patterns in asynchronous computer conferencing environments. *Computers and Education, 49*, 1258-1271*.

Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers and Education, 41*, 397-420.

Holliman, R., & Scanlon, E. (2006). Investigating cooperation and collaboration in near synchronous computer mediated conference. *Computers and Education, 46*, 322-335*.

Hrastinski, S., & Keller, C. (2007). An examination of research approaches that underlie research on educational technology: A review from 2000 to 2004. *Journal of Educational Computing Research, 36*(2), 175-190.

Hwang, W.-Y., Chen, N.-S., & Hsu, R.-L. (2006). Development and evaluation of multimedia whiteboard system for improving mathematical problem solving. *Computers and Education, 46*, 105-121*.

Jacobs, N., & McFarlane, A. (2005). Conferences as learning communities: some early lessons in using 'back-channel' technologies at an academic conference-distributed intelligence or divided attention? *Journal of Computer Assisted Learning, 21*, 317-329*.

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: does it contribute to

successful computer-supported collaborative learning? *Computers and Education, 49*, 1037-1065*.
Jeong, H., & Hmelo-Silver, C. E. (2010). *Technology use in CSCL: A content meta-analysis.* Paper presented at the Hawaiian International Conference on System Science.
Kitsantas, A., & Chow, A. (2007). College students' perceived threat and preference for seeking help in traditional, distributed, and distance learning environments. *Computers and Education, 48*, 383-395*.
Lai, C.-Y., & Wu, C.-C. (2006). Using handhelds in a jigsaw cooperative learning environment. *Journal of Computer Assisted Learning, 22*, 284-297*.
Lee, Y.-J. (2005). VisSearch: A collaborative Web searching environment. *Computers and Education, 44*, 423-439*.
Lehtinen, E., Hakkarainen, K., Lipponen, L., Rahikainen, M., & Muukkonen, H. (1999). Computer-supported collaborative learning: A review. Unpublished manuscript.
Li, S. C., Law, N., & Lui, K. F. A. (2006). Cognitive perturbation through dynamic modelling: a pedagogical approach to conceptual change in science. *Journal of Computer Assisted Learning, 22*, 405-422*.
Lim, C. P., & Barnes, S. (2005). A collective case study of the use of ICT in economics courses: A sociocultural approach. *The Journal of the Learning Sciences, 14*(4), 489-526*.
Lopez-Morteo, G., & Lopez, G. (2007). Computer support for learning mathematics: A learning environment based on recreational learning objects. *Computers and Education, 48*, 618-641*.
Mercier, J., & Frederiksen, C. H. (2007). Individual differences in graduate students' help-seeking process in using a computer coach in problem-based learning. *Learning and Instruction, 17*, 184-203*.
Miyake, N. (2006). Computer supported collaborative learning. In R. Andrew & C. Haythornwaite (Eds.), *Sage handbook of e-learning research.* London, UK: Sage.
Naidu, S., & Jarvela, S. (2006). Analyzing CMC for what? *Computers and Education, 46*, 96-103.
Nicol, D. J., & MacLeod, I. A. (2005). Using a shared workplace and wireless laptops to improve collaborative project learning in an engineering design class. *Computers and Education, 44*, 459-475*.
Redondo, M. A., Bravo, C., Ortega, M., & Verdejo, M. F. (2007). Providing adaptation and guidance for design learning by problem solving: The design planning approach in DomoSim-TPC environment. *Computers and Education, 48*, 642-657*.
Schellens, T., Van Keer, H., De Wever, B., & Valcke, M. (2007). Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? *International Journal of Computer-Supported Collaborative Learning, 2*, 225-246*.
Schwabe, G., & Goth, C. (2005). Mobile learning with a mobile game: design and motivational effects. *Journal of Computer Assisted Learning, 21*, 204-216*.
Stahl, G., Koschmann, T., & Suthers, D. D. (2006). Computer-supported collaborative learning: A historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences.* New York: Cambridge University Press.
Sung, Y.-T., Chang, K.-E., Chiou, S.-K., & Hou, H.-T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers and Education, 45*, 187-202*.
Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *Computer-Supported Collaborative Learning, 1*, 315-337.
Tseng, S.-C., & Tsai, C.-C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers and Education, 49*, 1161-1174*.
Veermans, V., & Cesareni, D. (2005). The nature of the discourse in web-based collaborative learning environments: Case studies from four different countries. *Computers and Education, 45*, 316-336*.
Vizcaino, A. (2005). A simulated student can improve collaborative learning. *International Journal of Artificial Intelligence in Education, 15*, 3-40*.
Wecker, C., Kohnle, C., & Fischer, F. (2007). Computer literacy and inquiry learning: when geeks learn less. *Journal of Computer Assisted Learning, 23*, 133-144*.
Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: concepts and assessment. *Learning and Instruction, 17*, 416-426.
White, T. (2006). Code talk: Student discourse and participation with networked handhelds. *International Journal of Computer-Supported Collaborative Learning, 1*, 359-382*.
Yang, Z., & Liu, Q. (2007). Research and development of web-based virtual online classroom. *Computers and Education, 48*, 171-184*.

## Acknowledgments

This research was in part funded by the National Research Foundation of Korea (NRF) (Grant No. 2009-0068919) awarded to the first author and a Rutgers University Faculty Research Council grant to the second author.

# A Visualization of Group Cognition: Semantic Network Analysis of A CSCL Community

Li Sha, The University of Hong Kong, Pokfulam Road, Hong Kong, lishahn@yahoo.com
Christopher Teplovs, Ontario Institute for Studies in Education, University of Toronto, Canada, chris.teplovs@utoronto.ca
Jan van Aalst, The University of Hong Kong, Pokfulam Road, Hong Kong, vanaalst@hku.hk

**Abstract:** This paper reports our progress in using the Knowledge Space Visualizer (KSV) as a tool for formative assessment of online discourse. Whereas social network analysis has been used in research on computer-supported collaborative learning, it only examines the social structure of discourse participants, and does not provide information about the content of the discourse. We discuss two types of networks as they relate to online discourse: *structural* and *semantic*. The initial findings indicate that the KSV can be used to visualize a Knowledge Forum database, and can provide a fine-grained semantic analysis that may enable teachers and students to locate the key ideas around which collective learning may takes place.

## Introduction

Sawyer (2006) identifies knowledge building as one of five major educational models produced by the learning sciences. As an educational framework, it aims to make the processes by which experts create new knowledge more prominent and feasible in schools (Bereiter & Scardamalia, 1993; Scardamalia & Bereiter, 2006). One of the most important features of the model is that students' effort is directed at advancing the collective knowledge in a community (Scardamalia, 2002). Students are not just attempting to understand things for themselves, but aim to add something new to what is known in the community. In most implementations in schools, students use Knowledge Forum®, a software environment specifically designed to support knowledge building, to share and collaboratively improve and synthesize ideas (Scardamalia, 2003). Knowledge Forum also provides a trace of how the community's ideas develop over time.

Although considerable progress has been made in the last decade to develop innovative classroom practices based on the knowledge-creation model in a variety of school subjects (see www.ikit.org and kbtn.cite.hku.hk), advances in assessment of electronic discourse have lagged. We believe that the development of a suite of assessment tools is urgent, particularly tools that teachers and students can use to self-assess and reflect on the nature of their work on Knowledge Forum. Previous studies have used content analysis to examine a variety of issues including knowledge advancement (Hakkarainen, Lipponen, & Järvelä, 2002; van Aalst, 2009; Zhang, Scardamalia, Lamon, Messina, & Reeve, 2007), but these methods are too labor-intensive to inform how students can improve their own knowledge-creation efforts. Analysis tools have been available for the last decade which students and teachers can use to examine participation patterns, such as note writing and reading and social network analysis (Burtis, 1998; de Laat, Lally, & Lipponen, 2007). While these techniques provide information about the social structure of students' collaborative discourse, they reveal little about how ideas are developing. They do not, for example, provide insight into questions such as "are the ideas of students becoming more coherent with one another other over time?" Computer-assisted visualizations or representations that capture both social and semantic features of the online discourse appear to hold much promise tools for self-assessment and reflection on knowledge-creation efforts at the community level.

The goal of this paper was to consider the nature of such representations and to explore the use of one visualization and assessment tool, the Knowledge Space Visualizer (KSV; Teplovs and Scardamalia, 2007; Teplovs, 2010). This work is part of a larger project, which will study the formative assessment *practices* by teachers and students, and will seek to develop a theory of formative assessment that uses data embedded in online environments like Knowledge Forum. However, in the present paper we report our initial exploration into the nature of the representations and what they reveal about the extent to which knowledge creation was occurring.

## Background

### Formative Assessment

In various forms, assessment drives educational practice (Biggs, 1996). It is therefore important for innovative educational approaches from the learning sciences to demonstrate that quality outcomes are obtained. However, it is equally important to understand how collaborative processes contribute to such outcomes. For example, how do we know whether discourse in Knowledge Forum is likely to lead to collective knowledge advances? How does new knowledge diffuse in a database?

We use the concept of *formative assessment* to frame our work: assessment that is used by students and the teacher to enhance learning, while the learning process still is in progress (Scriven, 1967). Interest in formative assessment received a new boost after the major review by Black and Wiliam (1998), who showed substantial positive impacts of formative assessment on learning. However, these practices seem to focus on such the provision of feedback on student work (e.g., tests and projects) and in-class questioning. Several authors have criticized the theoretical underpinnings of formative assessment (Perrenoud, 1998; Taras, 2005), and Black and Wiliam (2009) have recently proposed a theory of formative assessment that refers to self-regulated learning and situates formative assessment in an overall pedagogical framework. Most authors treat formative assessment as something that stands apart from the learning process. We rather consider it as a form of *inquiry* and as part of knowledge creation itself. Regardless of whether students are assessing their own work or whether they are advancing community knowledge by, for example, determining if forest fires are caused by human activities, the underlying process is the same: they use available evidence to ask questions and test hypotheses. This perspective requires that students have data about their own inquiry, which they can use to reflect on their progress and plan future actions. Such data are available but very complex. Students and teachers need sufficiently simple representations that can inform their reflection.

## Latent Semantic Analysis and the KSV

In the last decade, much attention has been given to "assessing" information on the Web to improve the performance of search engines. One example is Google's use of the PageRank algorithm to analyze the structure of the Web to identify the most influential Web pages, in combination of an index of the relevance of a Web page to a query (Brin & Page, 1998; Maslov & Redner, 2008). Latent Semantic Analysis (LSA) is another approach to the problem that uncovers the underlying semantic structure (meanings) of a network of documents (Landauer, 2007).

LSA is a *vector space* approach. One constructs a matrix (a table) in which the columns are types documents (e.g., books, chapters, essays, paragraphs, or computer notes) and the rows different terms (words or phrases); the entries are the numbers of occurrences of the terms in each of the documents. This matrix generally has a very large number of dimensions (e.g. when the "documents" are all the novels written in English or the entire Web), so mathematical techniques are used to produce a matrix of more manageable dimensions that is a reasonable representation of the whole vector space. The most common technique is Singular Value Decomposition (SVD). To make an assessment, one then constructs and compares vectors in the SVD. Vectors can contain information about the terms in a document, about the documents that use a specific term, or a combination. Geometrically, vectors are "similar" if they can be said to point in the same general direction, and if their lengths are comparable. For example, if the cosine of the angle between two vectors is 0.866 (in which case the corresponding angle is 30 degrees), the vectors can be said to point in the same general direction; if the length of one is 1.45 and the other 1.52 (rather than, say, 4.45 and 0.15) that would provide another indication of their similarity.

The Knowledge Space Visualizer (KSV; Fujita & Teplovs, 2009; Teplovs, 2010) is a Java-based tool that was developed to visualize networks of Knowledge Forum notes. It uses visual representations as well as quantitative network metrics to characterize idea-based networks. Exhaustive similarity measures between notes are recorded as latent semantic links between notes. These links, and the explicit semantic links afforded through referencing, rising-above and building-on functionality of Knowledge Forum, are then made available to the KSV. The KSV is capable of recreating the two-dimensional representation of collections of notes in Knowledge Forum, but it also provides computer-assisted positioning algorithms to facilitate the visualization of networks of notes.

## Collective Knowledge Advancement

According to Bereiter and Scardamalia's knowledge creation theory (Bereiter, 2002; Scardamalia & Bereiter, 2006), students in a knowledge creation classroom work progressively and collaboratively on a number of shared topics, and their collaborative inquiries lead to the advancement of individual and collective knowledge. Knowledge advancement is treated as a community rather than individual achievement (Scardamalia & Bereiter, 2006). Collective knowledge resides in conceptual artifacts (ideas and theories) in the community's discourse rather than inside individual minds. Also, according to Stahl's (2006) notion of group cognition, collective knowledge is interactively achieved in discourse and cannot be simply attributable to any particular individual mind.

These ideas suggest that any theoretical exploration, or practical assessment, of the advancement of community knowledge depend on the degree to which researchers and teachers are able to accurately describe and evaluate the content of discourses occurring in a collaborative knowledge creation community. However, despite emphasis and progress in developing collaborative inquiry in CSCL research, little attention has been given to the assessment of collective learning (van Aalst & Chan, 2007). Thus, little is known about the nature of collective knowledge and group cognition, and in practice teachers are still limited in their ability to assess how collective knowledge actually is formulated and advanced, although CSCL researchers and classroom teachers have put great efforts to design various technology-enhanced collaborative learning environments.

## Network analyses of CSCL

There are at least two significant factors underlying the above issues. First, research on CSCL often epistemologically focuses on individual learning outcomes rather than collective knowledge growth (Stahl, Koschmann, & Suthers, 2006). Perhaps this problem can be attributed to the influence of cognitivism, which stresses individual learning and personal cognition. Second, methodological alignment with the social nature of collaborative learning has led many researchers to examine interactions and participation patterns, using social network analysis (SNA; de Laat, Lally, Lipponen & Simons, 2007; Cho, Gay, Davidson, & Ingraffea, 2007). SNA considers a CSCL community as a *network* in which individual learners are represented as nodes, and the relationships between learners are represented as edges connecting those nodes. The edges indicate collaborative actions such as note reading, building-on, and referencing other notes. Current studies on the application of SNA to CSCL research converge at two basic aspects: network properties of a CSCL community, and participatory characteristics of individual learners. The first of these is mainly concerned with the issues such as network density of a collaborative learning community, and emergent cohesive cliques. The second aspect seeks answers to the following questions (de Laat, et al, 2007): Who is involved with the collaborative learning task? Who are the active participants? And who is participating peripherally? Current SNA studies rely primarily on mathematical computations of the frequency of ties between learners, and ignore the *content* of artifacts notes and the connections in meaning between the notes (Stahl, 2006).

For formative assessment, it is necessary to locate and describe knowledge creation in collaborative discourse. In network models of cognition, knowledge is represented as a network in long-term memory, in which nodes correspond to the cognitive units (in the form of concepts or schema), and the relations between the cognitive units are represented as links (Bruning, Schraw, & Norby, 2004). By analogy, group cognition can be defined as a network of the conceptual artifacts collectively created in a collaborative learning process. As information units, conceptual artifacts (Bereiter, 2002; Scardamalia & Bereiter, 2006) are the objects of collaborative work on knowledge creation, which take the form of notes created and posted by individual learners in Knowledge Forum. This implies that collective knowledge and group cognition can be studied and assessed from the perspective of network analysis. Therefore, we differentiate two types of networks existing in a CSCL community: *networks of people* and *networks of notes*. The former can serve to assess the interactive and participatory patterns of students, the latter can function as a representation of collective knowledge and group cognition.

However, as noted earlier, most research has concentrated on the patterns of social interaction between students rather than networks of conceptual artifacts. Of course, one may argue that a social network of people in a CSCL community operationally corresponds to a profile of the network of notes created by those people. Social networks of people only capture at most two features of notes – authorship or readership. Most SNA studies in CSCL examine patterns of social interaction between the CSCL participants, rather than the relation between students' ideas embedded in the notes from the perspective of collective knowledge and group cognition. Therefore, comprehensive formative assessments of collaborative learning must entail a multi-faceted network analysis of notes generated in a CSCL discourse. Stahl (2006) pointed out that the meaning of the group-level constructs such as group cognition constitutes a network of semantic references within the group interaction, and collaborative learning can be viewed as the interactive construction of this referential network. In this sense, collective knowledge and group cognition produce a semantic network of conceptual artifacts (notes), that is, a network system based on shared meaning (Doerfel & Barnett, 1999).

We may ask what a semantic network of notes looks like, how researchers or classroom teachers can identify the semantic networks of notes that emerge in a CSCL process, and what is the role of semantic network of notes in assessing collective knowledge advancement and group cognition.

This paper introduces our preliminary attempt to address these issues related to formative assessments of CSCL. Taken together, the above analyses of the distinction between the network of people and network of notes suggest that networks of notes can be established in two manners. Furthermore, we propose two sub-types of network of notes: *structural (physical) network* and *semantic network.* A structural network of notes or people refers to a visible network of notes or people whereby the linkage between notes or people is physically established by any one of the collaboration operations, such as note reading, building-on, and referencing in a Knowledge Forum database. An inquiry thread (Zhang, et al., 2007), defined as a series of notes that share a problem and constitute a conceptual stream in a collaborative inquiry, manifests the limitations of the studies on the spatial threads of notes, as well as structural network of notes in understanding the formation and advancement of collective knowledge and group cognition in a collaborative inquiry discourse. An inquiry thread represents a semantic connection between notes by content analysis of each individual note, namely, and therefore is a semantic network of notes from a network analysis perspective. Similarly, Manca, Delfino, and Mazzoni (2009) introduced a semantic coding scheme aiming to obtain a relatively complete picture of social interactions among people in a web forum that the structural (physical) network construct cannot provide. However, these methods require manual coding of note content, which is very labor-intensive. Because of the two considerable drawbacks—the amount of labor involved and their reliability—teachers may hesitate to use them for formative assessment. Our study is an attempt to meet this challenge, and uses the KSV, an innovative visualization and assessment tool of collective knowledge, to automatically identify the emergent semantic networks of notes under a pre-set condition with reference to semantic closeness between them, and the as structural networks of notes in a CSCL discourse.

## Research Context and Methods

The KSV was used to study part of a Knowledge Forum database created by a class of 41 Form 4 (Grade 10) students taking physics at a secondary school in Hong Kong. The database had approximately 880 notes, which were distributed over several views (shared workspaces) in Knowledge Forum. The teacher had several years of prior experience with Knowledge Forum, and was attempting to align her classroom teaching more with the knowledge-creation model, particularly by encouraging student-to-student classroom talk. The students' presentations and the talk that followed appeared to break with the more typical classroom discourse in Asian classrooms in which students do not ask many questions (Li, 2009). The curriculum consisted of three units of study: heat and temperature (3 months), mechanics (6.5 months), and waves (1.5 months).

In this classroom it was difficult to integrate the use of Knowledge Forum on a daily basis. Rather, it was used during specific periods to follow up classroom learning and to support the two projects. For example, during the heat and temperature unit, the class as a whole discussed thermal questions such as how to keep a drink warm as long as possible (Linn & Hsi, 2000) and why people lived in igloos, which are made from ice and would intuitively seem cold. Students also used Knowledge Forum to discuss their solar cooker designs, and in this, they were asked to refer to physics concepts and phenomena such as the greenhouse effect and methods for focusing energy from the sun on a beaker of water at the center of their solar cooker. During the mechanics unit, students similarly used Knowledge Forum to discuss puzzlements about Newton's laws and their analysis of motion on rides at an amusement park. Although the amount of work in Knowledge Forum was not as extensive as in some published studies, the computer notes that were generally focused on the tasks at hand and contained explanations in which students used relevant science ideas, although a substantial number of notes dealt with the logistics of projects.

## Results

The Knowledge Forum database had several comprehensive views for two units of study: heat and temperature, and mechanics. Each unit had several sub-topics. For example, the students studied mechanics by discussing how the cable car works in the view called 'Cable Car'. This view was spatially separated from the 'Mechanics' view.

### Two Types of Network of Notes

Two kinds of network were generated with the KSV. Figure 1 shows structural networks, and Figure 2 semantic networks. The directional lines in the structural network represent links between notes by note building-on. The lines in the semantic networks in Figure 2 are non-directional, suggesting that the notes are linked to one another in *meaning*, which is reciprocal. The color of node denotes a participant in Knowledge Forum (electronic version of the paper).

Mathematically, semantic closeness between notes in KSV is quantified as the cosine between vectors representing their content. High cosine values denote semantically similar note content (Teplovs & Scardamalia, 2007). The KSV allows us to set a link visibility threshold. When the cosine between any two notes exceeds this threshold, the KSV joins the notes with an edge (shown in Figure 2). The lower the threshold is, the more semantic links between notes in the network are displayed. In this study, the semantic coefficient was set at 0.75. Semantic links between notes are invisible in Figure 1 because KSV allows users to choose either of the two types of network.

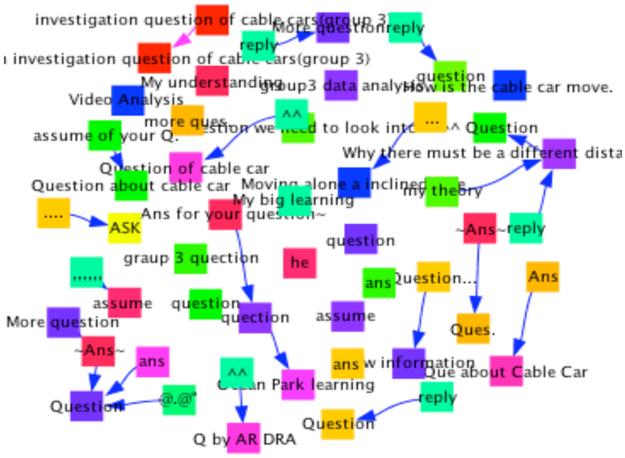



Figure 1. Obtained structural network in the Cable_car view

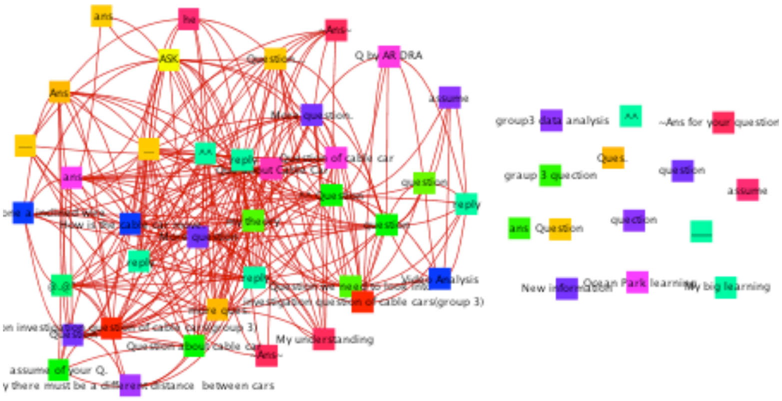


Figure 2. Obtained semantic network in the Cable_car view

Visually, the number of lines in the structural network is obviously much less than that in the semantic network, implying that many notes are linked to one another semantically but not structurally. This finding can be quantitatively reinforced by the contrast of the network properties between them. The Cable_car view contains 52 notes, 2391 words in total. The average size of note (i.e., words/notes) is 45 words. There are 22 lines (physical note links) in the structural network, which correspond to 22 physical actions (i.e., note building-on or referencing) that took place while the students collaboratively worked on that topic in Knowledge Forum. There are 281 links in the semantic network. In SNA, degree refers to the number of lines incident with each node in a graph; network density is defined as the ratio of the actual number of lines to the maximum possible lines, with a value falling within the range between 0 and 1 (Wasserman & Faust, 1994). The density of the semantic network is .212; the density of the structural network is .017, much less than the former. This finding implies that for a collection of notes in the Cable Car view, its semantic network is much cohesive than the structural network. The notes that are not linked to one another by a physical collaborative learning action are not necessarily disconnected with each other semantically.

## Visualization of Group Cognition

In the design of a KF database, the notes embedded within a view often converge at the same discussion topic. For example, the Heat view was intended to be a workspace where students collaboratively learned the concept of heat. In practice, a classroom teacher may want to examine how student online discussions in Knowledge Forum unfolded around the shared topic.

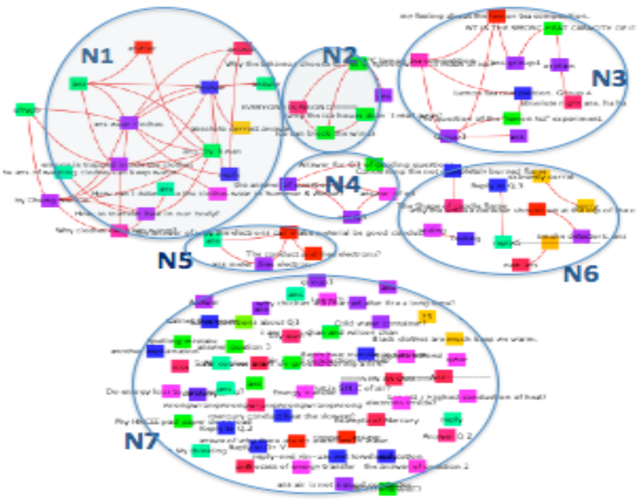



Figure 3. Obtained semantic networks within the Heat view

Table 2: The main overlapping words in each sub-semantic network

| Sub-networks | Main overlapping words |
|---|---|
| N1 | clothes (24) body (17) air (16) energy (14) warm (10) feel (6) transfer (6) cold (5) theory (5) trap (5) |
| N2 | ice (11) live (7) house (4), outside (3) |
| N3 | ice (16) energy (14) lemon (12) tea (11) loss (9) heat (8) specific (7) water (6) |
| N4 | heat (10) copper (9) water (7) specific (6) energy (5) capacity (5) conductor (3) |
| N5 | free (10) conduct (5) electrons (5) electron (5) metal (4) energy (4) shell (4) atom (3) heat (3) |
| N6 | air (17) copper (9) water (9) gravity (7) flame (6) electrons (6) |
| N7 | theory (33) air (23) energy (22) heat (20) liquid (13) ice (12) water (11) density (11 |

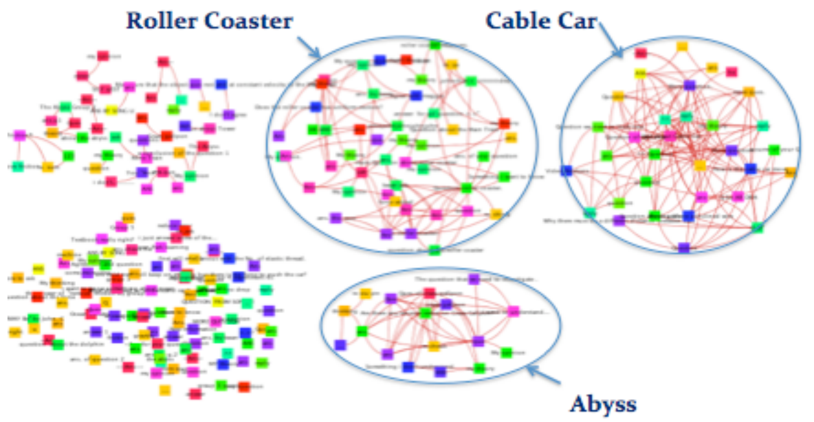



Figure 4. Sub-topics in the semantic network of notes in the Mechanics view

Figure 3 shows that the KSV revealed that there were seven clusters of notes within the general semantic network of the Heat view. These clusters can also be viewed as semantic sub-networks. Table 2 shows the major theme for each semantic sub-network. For example, the students who wrote the notes within N1 primarily were trying to understand the concept of Heat by connecting it to the daily life phenomena such as clothing from the perspective of heat transfer. Shared meaning of the notes in N3 seems to converge on heat transfer in liquids such as lemon tea, and water. Similar findings can be seen from the semantic network of another view in the same KF database shown in Figure 4 (Mechanics). Three themes in the view Mechanics can be identified by latent semantic analysis: Roller Coaster, Cable Car, and Abyss. Similarly, this finding suggests that students collectively worked on several sub-topics under the Mechanics view.

We conducted an inquiry thread analysis (Zhang et al., 2007) on the Heat view, and found seven inquiry threads (i.e., principal problems). Four themes identified in the inquiry thread analysis correspond to the first five semantic networks (i.e., sub-discussion topics) shown in Table 2. This provides some preliminary evidence that the KSV can reliably uncover sub-discussion topics that emerge in the discourse of collaborative knowledge creation.

## Conclusion

This study demonstrates a combination of computer-assisted content analysis and network analysis for formative assessments of CSCL, which is aligned with the appeal for developing innovative methodologies for analyzing participation and discourse processes in CSCL (Lipponen, Rahikainen, & Hakkarainen, 2002).

We can draw at least the following tentative conclusions. First, semantic network analysis offers an alternative approach for examining the relationship among the computer notes in a CSCL community in terms of shared meaning making, other than the traditional network analyses that mainly concentrate on the participation patterns of students. This expands the scope of formative assessments of collaborative learning by empirically revealing and conceptualizing two kinds of network of notes (i.e., structural network vs. semantic network). Second, various sub-discussion topics uncovered by latent semantic analysis in the heat view reveals the complexity of collaborative learning dynamics. This finding implies that collective learning in this class unfolded by breaking down a general concept/theory (e.g., Heat, Mechanics) into a number of sub-discussion topics (i.e., inquiry threads). Third, the notes with a certain semantic closeness may illustrate the workplace whereby collective knowledge might be forged due to the shared meaning of those notes. The overlapping words in a semantic network to some degree facilitate the visualization of the key meanings of the conceptual artifacts of group cognition. For the purpose of formative assessment of collective learning, the use of the KSV as an assessment tool provides a fine-grained semantic analysis that may enable teachers and students to locate the key ideas around which collective learning may takes place. It was found in this study that identification of the sub-topics within a view is generally aligned with inquiry thread analysis (Zhang et al., 2007). However, thread analysis used in that study was time consuming, because the theme of each thread of notes (shared topic) required labor-intensively content analysis of each Knowledge Forum note. Fourth, many current studies in assessment of CSCL use a whole CSCL community as unit of analysis, which might be appropriate when research is to assess student participation patterns by SNA throughout a CSCL process. Our study suggests that individual views as unit of analysis be appropriate as well for assessing the processes of collaborative knowledge construction and creation.

Our preliminary analyses using the KSV have yielded some findings regarding the characteristics of semantic network and the potential for further exploring the role of this kind of emergent networks in assessing how knowledge creation proceeds collaboratively in a Knowledge Forum database. Thus far, we still know little about the role of these emergent semantic networks in formative assessments of collaborative knowledge creation. The present study is just starting point along this direction. We believe that in order to have efficient and productive formative assessments of CSCL, semantic network analysis should be adopted along with other types of analysis such as content analysis of the notes within a selected semantic network.

## References

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Chicago, IL: Open Court.

Biggs, J. (Ed.). (1996). *Testing: To educate or to select? Education in Hong Kong at the crossroads*. Hong Kong, China: Hong Kong Educational Publishing Company.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice, 5*, 7-74.

Black, P., & Wiliam, D. (2009). Developing the Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*, 107-117.

Burtis, P. J. (1998). *Analytic toolkit for Knowledge Forum*. Toronto, ON.

Cho, H., Gay, G., Davidson, B., & Ingraffea, A. (2007). Social networks, communication styles, and learning performance in a CSCL community. *Computers & Education, 49*(2), 309-329 de Laat, M., Lally, V., Lipponen, L., & Simons, R-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *Computer-Supported Collaborative Learning, 2*, 87-103.

Fujita, N. and Teplovs, C. (2009). *Automating the Analysis of Collaborative Discourse: Identifying Idea Clusters*. Paper presented at CSCL 2009.

Hakkarainen, K., Lipponen, L., & Järvelä, S. (2002). Epistemology of inquiry and computer-supported collaborative learning. In T. Koschmann, R. Hall & N. Miyake (Eds.), *CSCL 2: Carrying forward the conversation* (pp. 11-41). Mahwah, NJ: Lawrence Erlbaum Associates.

Landauer, T. K. (2007). LSA as a theory of meaning. In T.K. Landauer, D.S. McNamara, S. Dennis & W. Kintsch (Eds.). *Handbook of Latent Semantic Analysis* (pp. 3-34). Mahwah, N.J. : Lawrence Erlbaum Associates.

Li, J. (2009). Learning to self-perfect: Chinese beliefs about learning. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: Changing contexts, changing education* (pp. 35-69). Hong Kong: Springer/Comparative Education Research Centre.

Linn, M. C., & Hsi, S. (2000). *Computers- teachers, peers: Science learning partners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lipponen, L., Rahikainen, M., & Hakkarainen, K. (2002). Effective participation and discourse through a computer network: Investigating elementary students' computer supported interaction. *Journal of Educational Computing Research, 27*(4), 355–384.

Manca, S., Delfino, M., & Mazzoni, E. (2009). Coding Procedures to Analyse Interaction Patterns in Educational Web Forums. *Journal of Computer Assisted Learning, 25*(2), 189-200.

Maslov, S., & Redner, S. (2008). Promise and pitfalls of extending Google's PageRank algorithm to citation networks. *The Journal of Neuroscience, 28*, 11103-11105.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy & Practice, 5*(1), 85-102.

Sawyer, R. K. (2006). *The Cambridge handbook of the learning sciences*. New York, NY: Cambridge University Press.

Scardamalia, M. (2003). Knowledge building environments: Extending the limits of the possible in education and knowledge work. In A. DiStefano, K. E. Rudestam & R. Silverman (Eds.), *Encyclopedia of distributed learning* (pp. 269-272). Thousand Oaks, CA: Sage Publications.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Chicago: Open Court.

Scardamalia, M. & Bereiter, C., (2006). Knowledge building: Theory, pedagogy and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–115). New York: Cambridge University Press.

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne & M. Scriven (Eds.), *Perspectives on curriculum evaluation*. Chicago, IL.: Rand McNally and Co.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Taras, M. (2005). Assessment - Summative and formative - Some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466-478.

Teplovs, C. and Scardamalia, M. (2007). *Visualizations for knowledge building assessment*. Paper presented at the AgileViz workshop, CSCL 2007.

Teplovs, C. (2010). *Visualization of knowledge spaces to enable concurrent, embedded and transformative input to knowledge building processes*. Unpublished doctoral dissertation. University of Toronto.

van Aalst, J., & Chan, K.K.C. (2007). Student-directed assessment of knowledge building using electronic portfolio. *The Journal of Learning Sciences, 16*, 175-220.

van Aalst, J. (2009). Distinguishing knowledge sharing, construction, and creation discourses. *International Journal of Computer-Supported Collaborative Learning, 4*, 259-288.

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-Cognitive Dynamics of Knowledge Building in the Work of 9- and 10-Year-Olds. *Educational Technology Research and Development, 55*, 117-145.

## Acknowledgments

This research was supported by a grant from the Research Grants Council of Hong Kong to the third author (Grant No. HKU 752508H). We also thank the teacher and students whose Knowledge Forum database we analyzed.

# The Effects of Physical and Virtual Manipulatives on Students' Conceptual Learning About Pulleys

Elizabeth Gire, Adrian Carmichael, Jacquelyn J. Chini, Amy Rouinfar & Sanjay Rebello, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506
egire@phys.ksu.edu, adrianc@phys.ksu.edu, haynicz@phys.ksu.edu, amy.rouinfar@gmail.com, srebello@phys.ksu.edu
Garrett Smith and Sadhana Puntambekar, University of Wisconsin, 1025 West Johnson Street, Suite 785, Madison, WI 53705
gwsmith@wisc.edu, puntambekar@education.wisc.edu

**Abstract:** With computers becoming more ubiquitous in our daily lives and in our classrooms, questions of how students interact and learn with physical experiments and computer simulations are central in science education. We investigated how students' ideas about pulleys were influenced by the use of physical and virtual manipulatives. We found that there were advantages for each type of manipulative, and that virtual and physical manipulatives helped students develop correct understandings of different concepts. We also found that the order the manipulatives were used affected student learning, with students who used real pulleys before the simulation achieving higher scores on questions having to do with effort force, the distance the rope is pulled, and mechanical advantage.

## Introduction & Background

Laboratory experiments play a critical role in furthering scientists' understandings of how the universe works, and in light of this importance, it is no wonder that educators have historically placed high value on laboratory experiences in science classrooms. However, due to practical concerns of procuring laboratory equipment, safety concerns, and time constraints, computer simulated experiments are becoming an attractive alternative to laboratory experiments. In light of this trend, recent research in science education has explored whether computer simulations (virtual manipulatives) can be as effective for learning as experiments involving real objects (physical manipulatives) and researchers have begun looking at the circumstances in which these two alternatives may be best employed.

Finkelstein *et al.* (2005) investigated how physical versus virtual manipulatives supported students' learning about circuits. Students used either physical materials or simulations to examine combinations of resistors, build simple circuits, predict the behavior of specific elements and develop a method for measuring resistance. The simulations were similar to the set-up with physical materials, except that the simulations represented electron flow within the circuit, an aspect of the physical materials that cannot not directly be perceived. After these experiences, students who had used the virtual manipulatives were able to build physical circuits quicker than students who had previously used the physical manipulatives. In addition, the students in the virtual conditions were able to provide better explanations of circuit behavior and scored better on a related exam question. Therefore, Finkelstein *et al.* suggest properly designed simulations can be beneficial to student learning when applied in the appropriate contexts.

Triona, Klahr and Williams (2007) investigated how physical and virtual manipulatives support students' learning about the factors affecting how far a mouse trap car will travel. Students explored these factors by designing cars to be used for an experiment. Students used either physical or virtual manipulatives and were allowed to design either a certain number of cars or were allowed to design cars for a certain length of time, creating four treatment groups. All treatments were equally effective at increasing students' knowledge about causal factors for travel distance, supporting students' ability to design cars, and students' confidence in their knowledge. Based on these findings, the researchers suggest that simulations may be preferred due to their other pragmatic advantages.

Zacharia, Olympiou, & Papaevipidou (2008) studied physical and virtual manipulatives used in combination to learn about heat and temperature. Students in the control group used only physical manipulatives, while students in the experimental condition used physical manipulatives followed by virtual manipulatives. The researchers aimed to limit the differences between the physical and virtual manipulatives to speed of manipulation. On a conceptual test, students in the experimental group outperformed students in the control group. The researchers suggest this difference may be a result of virtual manipulatives being manipulated faster than physical manipulatives.

In a similar study (Zacharia & Constantinou, 2008), the researchers controlled for all differences between the physical and virtual conditions except for the mode in which experiments were performed. In particular, the simulations did not model any aspects of the phenomena that could not be perceived with the physical manipulatives (in contrast to Finkelstein, et al). In this case, the physical and virtual manipulatives equally supported students' conceptual understanding.

In our study, we also controlled for all conditions (curriculum, mode of instruction and resource capabilities) except for the mode of the activities (physical or virtual). Students spent approximately 30 minutes on each activity, although working with the real pulleys typically took a few minutes longer than working with the simulation. The design of this study replicates the study performed by Zacharia & Constantinou in a new domain (pulleys rather than heat & temperature). Furthermore, we not only looked at overall learning during these activities, but isolated particular concepts and looked at the effect of manipulative type and ordering of manipulatives on students' understandings of these concepts.

## Context and Data Collection

Students in a university-level conceptual physics lab performed two activities to learn about pulleys. One activity involved working with real pulleys (physical manipulatives) while the other activity involved an interactive computer simulation of pulleys (virtual manipulatives) (Figure 1). The activities are part of CoMPASS, a design-based curriculum that integrates concept maps and hypertext that students explore prior to performing physical or virtual experiments (Puntambekar, Stylianou & Hübscher, 2003 and Puntambekar & Stylianou, 2002). During each activity, students answered questions on a worksheet. These worksheet questions were the same for both activities. However, the temporal order of the activities was varied creating two treatment groups. Three sections (N=71) used the physical pulleys first (the Physical-Virtual treatment), while two sections (N=61) began with using the virtual pulleys (the Virtual-Physical treatment).

Students answered a set of conceptual assessment questions before the activities (pre-test), after the first activity (mid-test), and after the second activity (post-test). The assessment questions on the pre-, mid-, and post-tests were identical. The mid-test scores allowed for comparisons to be made between the effects of physical manipulatives (PM) and the effects of virtual manipulatives (VM) only, while mid-test and post-test scores indicated ordering effects.

The assessment contains 13 multiple-choice questions, with each question weighted equally in the total score. The assessment questions were developed locally to probe students' conceptual understanding of pulley concepts, including effort force, work, mechanical advantage, the distance the rope is pulled and the potential energy of the load. The assessment contained more questions about effort force and work concepts than other concepts because these are the most central to the topics of pulleys and the most applicable in other science topics. Figure 2 indicates the distribution of questions for each concept. A question was considered to be related to a concept when the concept is explicitly mentioned in the problem statement. For example, the question "If we ignore friction, what will require *less effort* (force) to lift a box to a height of 1 meter – using the pulley shown or lifting the box straight up?" is considered to be an effort force question.

A reliability analysis was conducted on effort force and work questions. For the effort force questions Cronbach's $\alpha$ = .70. For the work questions, Cronbach's $\alpha$ = .51. The lower reliability for the work questions may indicate that students have a harder time constructing correct understandings about the concept of work.

Open-ended worksheet questions were coded and analyzed using a phenomenographic approach (Marton, 1986). The conceptual assessments were analyzed statistically. For these assessments, categories of questions were created based on the physics concept probed by each question (as indicated explicitly in the question statement). The analysis included comparisons of overall scores and category scores that were made using a Repeated Measures Analysis of Variance with a between subjects factor of treatment type. P-values less than 0.05 were interpreted to indicate a statistically significant difference.

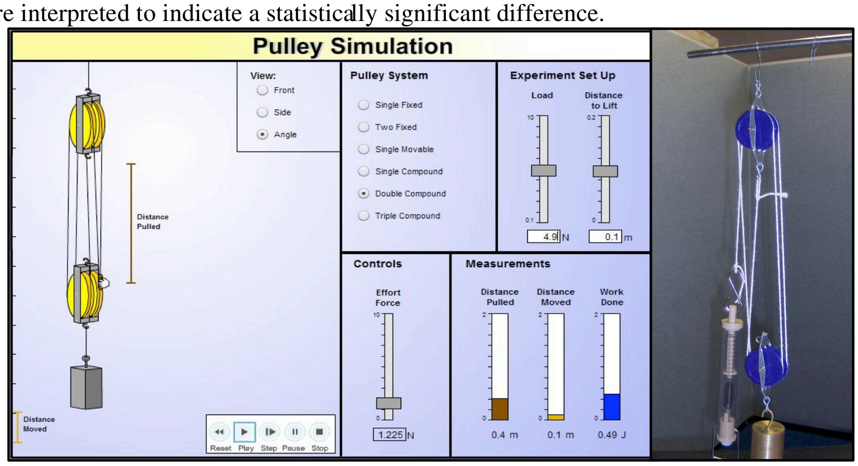



Figure 1. Virtual (left) and physical (right) manipulatives.

## Data Analysis

### Conceptual Assessment

The overall scores for the pre-, mid- and post-tests are shown in Table 1. Table 2 contains the results of a Repeated Measures Analysis of Variance for all the students. Mauchly's test indicated that the assumption of sphericity had been violated on all the comparisons made, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Table 2 also contains information of Mauchy's test and sphericity estimates.

Table 1: Overall pre-, mid-, and post-test scores on the conceptual assessment. Uncertainties are the standard error of the mean.

| Treatment | N | Pre-test % | Mid-test % | Post-test % |
|---|---|---|---|---|
| Physical-Virtual | 71 | **37 ± 2** | **58 ± 2** | **66 ± 3** |
| Virtual-Physical | 61 | **33 ± 2** | **60 ± 3** | **61 ± 3** |

Table 2: Mauchly's Test, Greenhouse-Gessier Estimates of Sphericity, and Repeated Measures ANOVA for Overall Score, Force Questions, and Work Questions

| Effect | Mauchly's Test | | Sphericity | Repeated Measures Analysis of Variance | |
|---|---|---|---|---|---|
| | $\chi^2(2)$ | p | ε | F | p |
| Total Score | 67.28 | <.001 | .71 | F(1.42,184.87) = 173.57 | <.001 |
| Total Score*Treatment | | | | F(1.42,184.87) = 2.33 | 0.12 |
| Effort Force Questions | 61.40 | <.001 | .73 | F(1.45,188.58) = 167.24 | <.001 |
| Effort Force*Treatment | | | | F(1.45,188.58) = 4.89 | 0.02 |
| Work Questions | 25.81 | <.001 | .85 | F(1.69, 220.09) = 27.69 | <.001 |
| Work*Treatment | | | | F(1.69, 220.09) = 15.28 | <.001 |

The Repeated Measures analysis shows that students' total scores changed significantly between tests. However, the insignificance of the interaction between total score and treatment condition indicates that the changes in scores for the two treatments were similar. The scores for the effort force questions also show a significant change between tests and a significant interaction between the effort force questions and the treatment conditions indicating that the changes in effort force score were different for the two conditions. Similarly, the scores for the work questions show a significant change between tests and a significant interaction between work score and treatment condition.

Table 3 shows the results of contrast comparisons for each effect described in Table 2. These comparisons help locate when significant changes in score occurred. Both treatment conditions resulted in a change of total score between the pre- and mid-test, but the scores of the treatment conditions changed differently between the mid- and post-test. Figure 2 shows plots of average score on each exam. The plot shows that students in the physical-virtual condition had a significantly greater change in score between the mid- and post-test. Therefore, the students who used the physical manipulatives first continued to progress on the conceptual assessment after the second activity while the students who used the virtual manipulatives for showed little further progression, although students in both treatment conditions benefitted from the activities on the whole.

On Effort Force questions, both conditions showed a change in score between the pre- and mid-test, but no change in score was observed between the mid- and post-test. The change in score for the treatment conditions was significantly different between the pre- and mid-test, with students who initially used the physical manipulatives showing a larger gain in Effort Force score than students who initially used the virtual manipulatives.

On the Work questions, an overall change in score was observed for between pre- and post-test. However, the changes were significantly different for the different treatments. Students in the virtual-physical condition show a large gain in Work score between pre- and mid-test while the students in the physical-virtual condition do not show a gain. However, this trend switches between the mid- and post-test, with the virtual-physical students showing no gain and the physical-virtual students showing a large gain. It seems that students show a large increase in score on work questions after they have used the pulley simulation.

For the other questions dealing with mechanical advantage, distance pulled, and potential energy, students in both conditions showed an increased of score between pre- and post-tests but no interaction effect was observed for these questions.

Table 3: Contrast Comparisons for Repeated Measures ANOVA

| Effect | Comparison | F(1,130) | p |
|---|---|---|---|
| Total Score | Pre-Mid | 170.94 | <.001 |
| | Mid-Post | 22.33 | <.001 |
| Total Score*Treatment | Pre-Mid | 1.71 | .19 |
| | Mid-Post | 12.04 | .001 |
| Effort Force Questions | Pre-Mid | 181.30 | <.001 |
| | Mid-Post | 1.74 | .19 |
| Effort Force*Treatment | Pre-Mid | 5.56 | .02 |
| | Mid-Post | <.001 | .98 |
| Work Questions | Pre-Mid | 10.83 | .001 |
| | Mid-Post | 26.27 | <.001 |
| Work*Treatment | Pre-Mid | 24.18 | <.001 |
| | Mid-Post | 28.14 | <.001 |

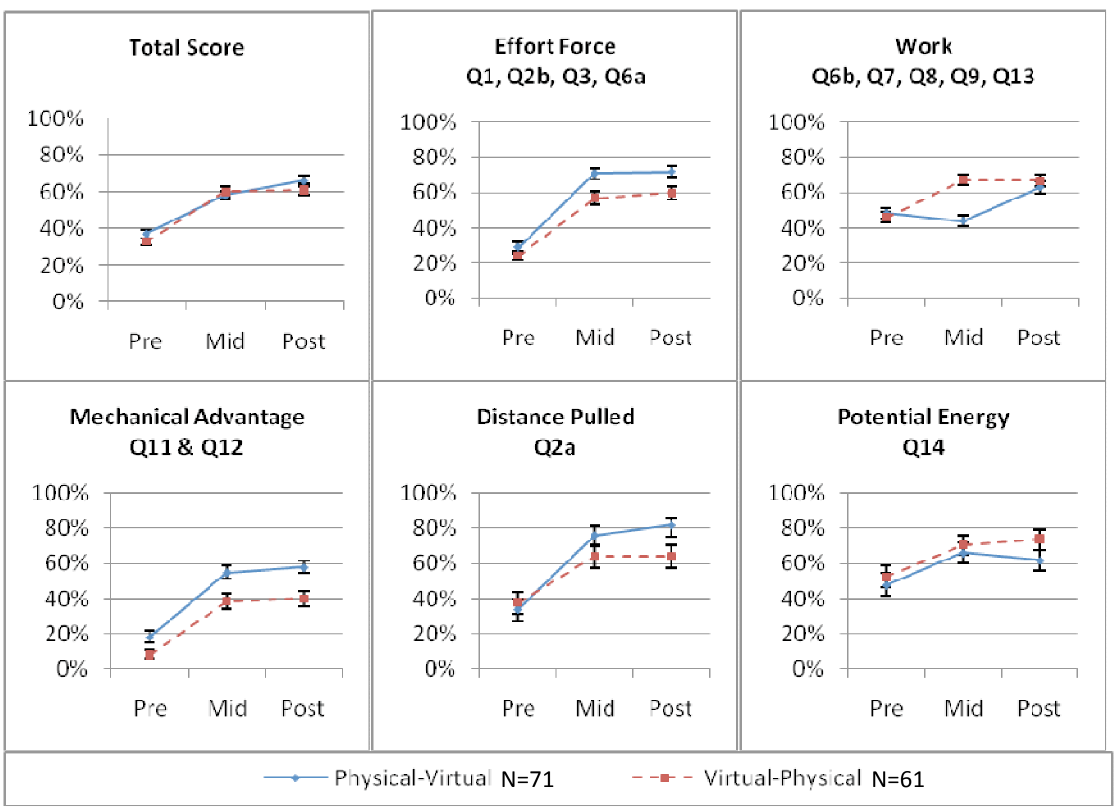
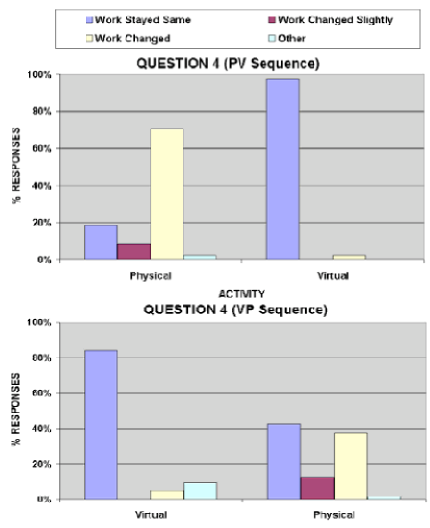



Figure 2. Average scores by category on the conceptual assessments. Error bars indicate standard error.

## Questions on Activity Worksheet

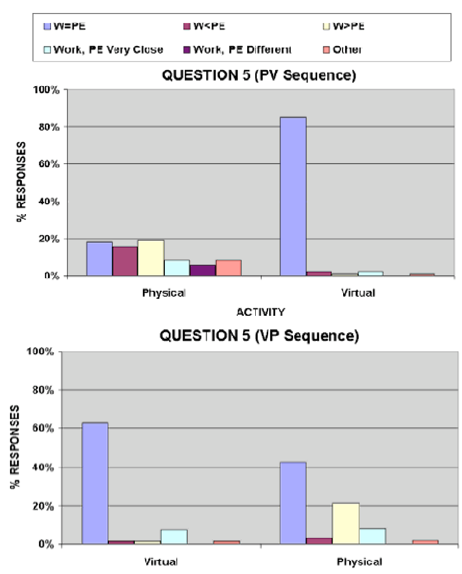
While doing the activities (with physical pulleys and virtual pulleys), students responded to questions on a worksheet. We report students' responses on some of the worksheet questions to aid interpreting the above assessment results. Question 4 on the worksheet had to do with the concept of work (Figure 3). Students in both treatments interpreted the data from the simulation as showing the work being the same for different pulley-setups. In contrast, students in the different treatments disagreed on how to interpret the data from the real pulley: students in the Physical-Virtual treatment said the work changed when you changed set-ups while students in the Virtual-Physical treatment were split, with nearly half of the students claiming the work stayed the same while the other half said the work changed across pulley set-ups.

A similar trend was seen on Question 5 on the worksheet (Figure 4). Question 5 had to do with comparing work and potential energy for a given pulley system. Students in both treatments interpreted the data from the simulation as showing the work was equal to the potential energy for a given pulley set-up. In contrast, students in the different treatments disagreed on how to interpret the data from the real pulley: students in the Physical-Virtual treatment did not come to a consensus about how the work was related to the potential energy while students in the Virtual-Physical treatment were more likely to say the work was equal to the potential energy.




Question WS4: "Based on your data, when you changed the pulley setup, how did it affect the work required to lift the object? Why do you think that is?"

Figure 3. Student responses to Question 4 on the activity worksheet.




Question WS5: "Based on your data, how does work compare to potential energy for a given pulley system? Why do you think that is?

Figure 4. Student responses to Question 5 on the activity worksheet.

## Discussion

The analysis of overall score indicates that there is no overall preferred manipulative for learning about pulleys; the students' mid-test scores (which isolate each type of manipulative) are the same. Looking at changes between the mid-test and post-test scores, however, indicates that there may be an ordering effect for using both types of manipulatives, in that the scores of students in the Physical-Virtual treatment continued to increase after the second activity, while the scores of students who used virtual pulleys first did not increase when the students switched to physical pulleys.

In looking at categories of questions, the data show that the different types of manipulative (physical or virtual) affected different pulley concepts differently. In the Effort Force category, although an increase in score was observed on mid-test for both treatments, students who used real pulleys scored higher. In contrast, students showed no increase in score on questions about work after using real pulleys, while students who used the pulley simulation showed significant improvement on work questions.

Additionally, the Effort Force category showed an ordering effect on category score, while the Work category did not. Students who used the physical pulleys first showed a greater increase in score between the pre- and mid-tests, while neither treatment group showed an increase in score between the mid- and post-tests. In contrast, students who used virtual pulleys more often answered questions about work correctly than students who only used physical pulleys, regardless of which order they encountered the manipulatives. In light of these data, it seems the ordering effect seen in the overall score may be explained by the ordering effect of questions having to do with effort force.

In short, students obtain a better understanding as measured by the conceptual assessment of the concept of work with the computer simulation and a better understanding of effort force with the real pulleys.

Why should the type of manipulative affect different pulley concepts differently? One possibility is that this result is due to a difference in the salience of the physics concepts. The concepts of effort force and distance pulled are more salient for real pulleys than with virtual pulleys. The real pulleys give students a kinesthetic experience with effort force and distance pulled; they feel the force they need to exert and how far their arms move in relation to the pulley, while students who use the virtual pulleys read-off values from a screen. On the other hand, work is less salient than effort force for real pulleys. Work depends on the combination of effort force and distance pulled, and this makes the concept of work more removed from the kinesthetic experience. In order to reason about work correctly, students need to coordinate the experiences that if less force is needed, the distance pulled is longer in exactly the right proportion so that work is constant across pulley set-ups (assuming friction is small enough to be neglected). Alternately, students might perceive the energy expended in lifting a load a certain distance with different pulley set-ups. Both of these types of reasoning are difficult to achieve through kinesthetic experience. However, all concepts are equally salient in the simulation – all quantities must be read-off the screen (Figure 1).

Capacity theories of attention suggest that people can only attend to a limited amount of information (Kahneman, 1973), and it is also known that attention can be influenced by the salience of cues, with high salience cues naturally attracting more attention (Denton & Kruschke, 2006). With physical manipulatives, effort force and distance pulled have a relatively high salience and probably dominate the attention of the learner, while less salient concepts, like work, receive less attention. This explanation is consistent with our data that Physical-Virtual students have relatively high scores on Effort Force questions and relatively low scores on Work questions. Students who use the virtual manipulatives probably divide their attention more evenly among the concepts due to their equivalence in salience, resulting in less relative attention to effort force, distance pulled and mechanical advantage concepts, resulting in lower scores in these categories and higher scores on work questions than students who used physical manipulatives. Furthermore, if the simulation is done first, the initial equivalence in salience among concepts may lessen the impact of the subsequent kinesthetic experience with the real pulleys, resulting in the continued suppression, or blocking (Kamin, 1968), observed in the Effort Force, Distance Pulled scores, and Mechanical Advantage categories.

Furthermore, the students in this study did not receive instruction on how to interpret their data in light of friction effects and experimental uncertainty. The data show that they did not interpret trends in their real pulley work data in a consistent way. This point is supported by the students' responses to Question 4 and Question 5 on the worksheet. Also, Physical-Virtual students had trouble reasoning about work in frictionless situations on the mid-test. However, after the virtual experiment, all students were much more successful in reasoning about work in frictionless environments, regardless of the order the activities.

This study suggests that curricula that include pulleys might ideally use both experiments with real pulleys and simulations. The data suggest that an ideal ordering might be to have students begin with real pulley experiments, focusing on effort force, distance pulled and mechanical advantage. Then students might perform a set of experiments with the simulation, reinforcing trends found with the real pulleys and exploring the concepts of work and potential energy in a frictionless simulated environment. Finally, students might finish with a set of experiments with real pulleys to explore how work and potential energy are related in a situation where friction is not negligible.

## Conclusion

Previous research has demonstrated that virtual manipulatives can be as effective as physical manipulatives in some circumstances. Our data extends this result to the domain of pulleys and addresses how manipulative types affect different pulley concepts. In looking at the effect of physical and virtual manipulatives on students' understandings of pulleys, we find that although there is no difference in overall score between types of manipulative, each manipulative has advantages for different pulley concepts: physical manipulatives better address the concepts of effort force, distance pulled and mechanical advantage, while virtual manipulatives better address the concept of work. The order the manipulatives are used by the students affected conceptual gains for the concepts of effort force, distance pulled and mechanical advantage. We suggest that these differences may be attributed to differences in concept salience.

## Endnotes

(1) The assessment and worksheets can be found online at http://www.phys.ksu.edu/personal/egire/Resources.html

## References

De Jong, T. & Van Joolingen W.R. (1998) Scientific Discovery Learning With Computer Simulations of Conceptual Domains, *Review of Educational Research*, 68, 179-201.

Denton, S.E. & Kruschke, J.K. (2006) Attention and salience in associative blocking. *Learning & Behavior,* 34(3), 285-304.

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Kohl, K. K., Podolefskey, N.S., et al. (2005). When learning about the real world is better done virtually: A study of substituting simulations for laboratory equipment. *Physical Review Special Topics- Physics Education Research*, *1*, 010103.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior: Aversive stimulation* (pp. 9-33). Coral Gables, FL: University of Miami Press.

Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching, 44*(1), 183-203.

Marton, F. (1986). Phenomenography- a research approach to investigating different understanding of reality. *Journal of Thought, 21*, 29-39.

Puntambekar, S., Stylianou, A., & Hübscher, R. (2003). Improving navigation and learning in hypertext environments with navigable concept maps. *Human-Computer Interaction, 18*, 395-428.

Puntambekar, S. & Stylianou, A. (2002). CoMPASS: Students' use of external representations in science learning. In P. Bell, R. Stevens & T. Satwicz (Eds.), *Keeping Learning Complex: The Proceedings of the Fifth International Conference of the Learning Sciences (ICLS)* (pp.352-358). Mahwah, NJ: Erlbaum.

Triona, L. M. & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction, 21*(2), 149-173.

Zacharia, Z. C. & Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the Physics by Inquiry curriculum: The case of undergraduate students' conceptual understanding of heat and temperature. *American Journal of Physics, 76*(4&5), 425-430.

Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. *Journal of Research in Science Teaching, 45*(9), 1021-1035.

## Acknowledgments

Foremost, we would like to thank the students and staff at CoMPASS Group, Wiconsin Center for Educational Research for their cooperation. We would also like to thank the students who participated in this study, particularly those who volunteered to be interviewed. Their cooperation and candor was critical to this work. Additionally, we'd like to thank the members of the Physics Education Research Group at Kansas State University for fruitful discussions and their thoughtful suggestions. This work was supported in part by the National Science Foundation under DUE grant DGE 0841414 and U.S. Department of Education IES Award R305A080507. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation or the U.S Department of Education.

# Children Learning Literate Practices in Spriting

Tara Rosenberger Shankar, 22 Freeman Street, Arlington MA 02474, tara@media.mit.edu

**Abstract:** The exploratory empirical work presented here suggests that children engage in conceptual processes through a new activity called *spriting* (Shankar 2005, 2006) previously thought of as late-emerging and dependent on writing. Spriting is a new technologically-supported composition process that does not require intermediate mediation with a written form. The output of spriting is a *talkument*. In this paper, results are presented from a design research study of the impact of spriting on the literacy development of children, ages 5-10, at two different schools. In particular, composite composing and editing actions in spriting presage a reversal of some trends seen in written composition, such as older children producing longer compositions. Furthermore, this paper describes early observations of children appearing to make certain literate habits, such as pausing to plan, explicit in spriting in ways they had not learned to do yet in writing. Results suggest that spriting could offer an alternative way to explore extended forms of literacy and habits of mind before a child is able to write.

## Introduction

Learning how to write has long been seen as a precursor to developing critical cognitive skills, typically lumped under the term "literacy". For example, children who cannot read or write are not expected to understand the difference between 'meaning' and 'saying' (e.g. Olson & Torrance, 1985), understand that compositions encode purpose and make things happen in the world (e.g. Bereiter & Scardamalia, 1987), or learn to recontextualize their oral language skills for a distant and critical audience (e.g. Purcell-Gates, 1991). This article explores the idea that children are much more capable of shaping their actions and communication in a literate manner than we have been able to observe given the literate tools available. There is some evidence that composition tools impact editing practices. Word processors can positively impact more experienced writers' revision practices (Haas, 1989) and even have a weak effect on children's writing as young as kindergarten (Jones, 1998). This work introduces a new technological support for composing in order to study the impact on children with emerging literacy skills.

In order to put this work in the broad frame intended, some new terms are required. First, I introduce a new practice called *spriting* (rhymes with writing), which refers to using one's own oral productions in a writerly process of composition, supported by new technology. Pragmatically, the spriting process goes something like this: children hold or wear a microphone that plugs directly in to the computer. Using custom built software called the SpriterWriter (see the interface presented in Figure 1), the child presses a record button and speaks for as long as s/he wants, and then presses a stop button to end the recording. While the child is recording, her/his speech productions appear as graphic representations in the software. The software automatically segments the speech productions into "sausages", based on spoken phrase automatic end-point detection. These smaller sausage units are independently editable and are intended to expedite review and editing. Through simple point and click interaction with the graphic representations, the child can *aud* the work (a correlate to reading in the aural mode), and apply editing functions (e.g. rearrange by drag and drop; cut, copy, and paste; delete; insert). Theoretically, the output of a spriting process could be an oral *talkument* (e.g. an audio file) or a written *document*, if the child would transcribe his or her own spriting or was supported by a technological process such as speech recognition. In practice, children rarely transcribed their spriting to writing, more often working very hard to realize their compositions goals as talkuments.

In short, spriting does not require children to make the additional leap to using a deeply different form of composition representation such as writing itself. Yet, spriting shares important properties with writing, such as permanence and shareability, and permits extended processes of planning and editing.

The introduction of spriting helps make a new distinction in literacy between alphabetic and higher-level skills. I refer to the knowledge and skills that bridge reading and writing, and auding and spriting as *literacy.* I borrow a term from Seymour Papert, *letteracy* (Papert, 1993), to refer to the mechanical and presentational skills specific to reading and writing, including the ability to write letters, spell, punctuate, and design a page. In spriting, I observed many parallels to written mechanics in this new compositional mode. Children were as concerned with *how* they said something—and frequently edited for these new "oral mechanics"—as they were concerned with *what* they said. I have dubbed these new material concerns, *prosodacy,* referring to acoustic-phonetic features of the voice, paralinguistic features of speech, and the age-old art of poetics, composing words pleasing to the ear. This new distinction between literacy and letteracy, pushed to a radical position, means that it is possible for someone to be highly literate but not know how to read or write.

This paper presents quantitative and qualitative empirical data and observations from a study done when I was a doctoral student at the MIT Media Laboratory that show how spriting enabled children, ages 5 to 10, to compose, edit and reflect on their compositions in ways that differ in marked ways from their emergent writing practices. While the full scope of the investigation was very broad, including results on editing, collaboration, vocabulary, and singing, and technology development (Shankar 2005), my purpose here is to focus only on some very specific results in the children's editing practices.

My approach in this paper is as follows. First I describe the exploratory research method used, the three populations of children who assisted this work, and the dataset generated. Secondly, I provide a quantitative overview of composite composition and editing moves to sketch some overall trends between age groups and SES, typical predictive indicators of literacy development. Thirdly, I discuss how the introduction of a new compositional form can produce new insights into the composition process generally, as it permits comparisons where few have existed before. To do so, I present two learning stories of children spriting and how they handled the new challenge of integrating moments of pause for thinking. Finally, I conclude with a discussion

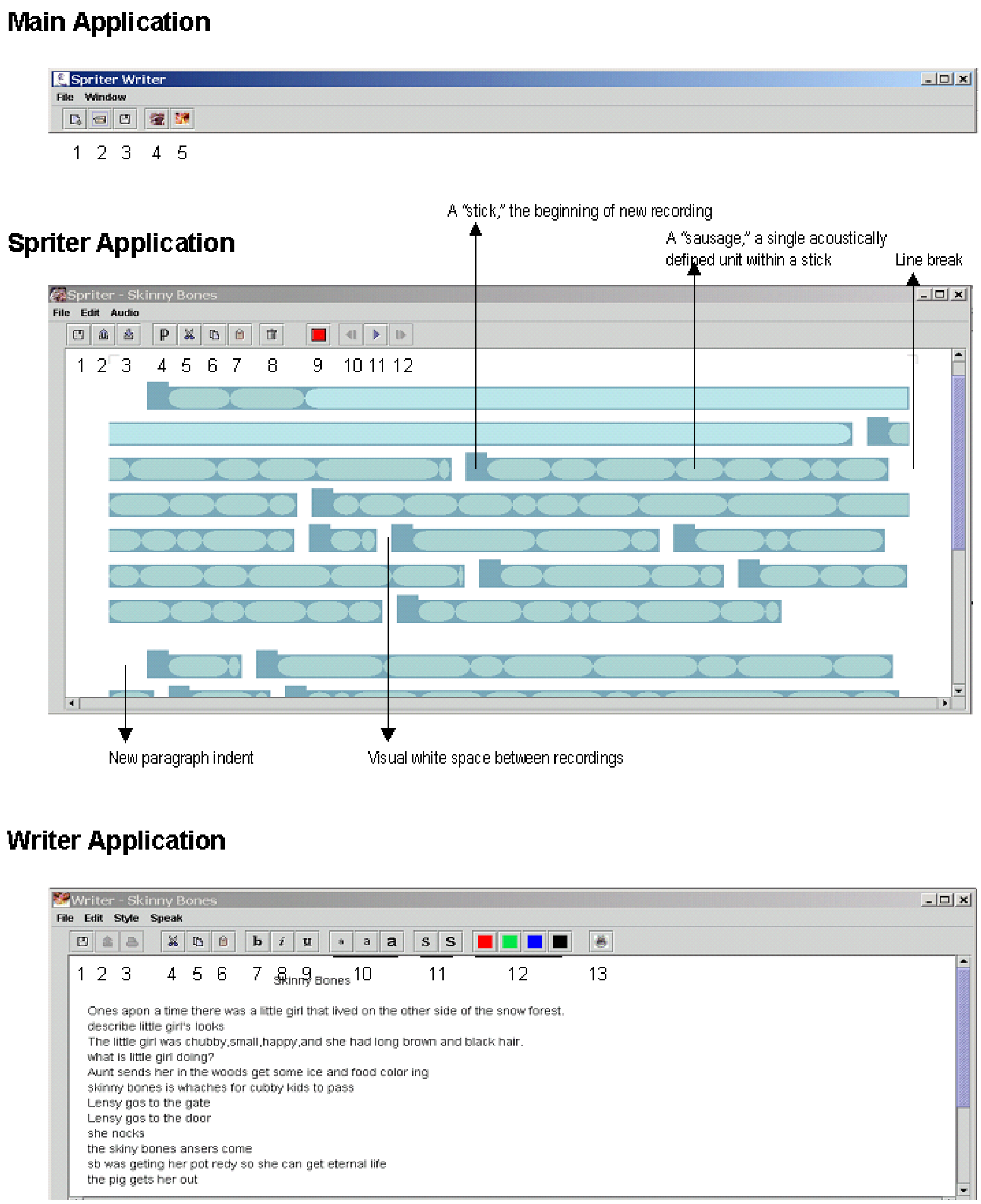



**Figure 1 The SpriterWriter interface.**

on implications for literacy learning.

## Methodology

I used a *design research* approach, inheriting from *design experiments*, which are also intended to evolve new learning technologies (Brown, 1992; Collins, 1992) and develop and refine new theories (diSessa & Cobb, 2004; Edelson, 2002) within real learning environments. Designing certain kinds of tools that challenge many dear cultural values, like writing, and require complex and sophisticated research, like spriting, might be seen as a process of evolving the ends of design as much as the means of design. While design work like this should prepare the way for many small kinds of experiments to follow, we must also allow time for design and cultural change to evolve contemporaneously. A design experiment approach was one way to allow the ends to evolve through the influence and needs of the learners themselves.

I performed initial exploratory and prototyping work with low-income adult learners in East Harlem, NYC, who were at elementary levels of letteracy learning. This research suggested that spriting can be a useful tool for supporting certain kinds of learning central to composition and editing. With particular questions and a spriting prototype, I undertook a thirteen week long study with children ages 5 to 10 at two elementary schools, here referred to as Umoja and Molière elementary schools, to explore how an improved spriting technology might contribute to their literacy and letteracy learning. I revised the software in cycles that ranged from a week to several weeks, according to my observations of its usability and the enthusiasm with which the children embraced it to do their work.

The design research experiments were conducted at two private elementary schools, both located in different urban-residential areas of a major city in the USA. The names of the schools and children referred to are pseudonyms. I call the first school Umoja Elementary, a private school founded upon African principals of community and culture (Nguzo Saba). The school functions as a one-room school house concept, not assigning children into grades per se. The students are children of color (Caribbean, African, African-American), low to middle SES, ages 5 to 10, and represented all the students in the school – not a select subset. The group ranged from 6 students when I began (referred to as the "older" Umoja students), and expanded to 11 students when the five/six year olds 'graduated up' (referred to as the 'younger' group) a few weeks later. The Umoja younger children were learning character formation rules at the time. While STEM activities were creatively realized in the Umoja curriculum, reading and writing activities were conventionally practiced with vocabulary words lists, fill-in-the-blank worksheets, some personal reflection writing prompts, and occasional stories read by the teacher. The children were invited to sprite once per week or more, and were able to decline if they didn't want to. For these children, choosing to sprite - whether or not they had the time, inspiration, and persistence to actually compose anything - must be understood as highly motivated and enthusiastic behavior.

The second school was Molière Elementary, a private French-English bilingual school. I offered an after school spriting club once a week, to which ten 3rd and 4th grade students responded (e.g. Irish, Russian, French-Canadian, French, and African families), a club which served to unintentionally self-select students from families with higher learning and research connections. The children were high SES, ranking in the top half of their classes in both French and English language exams; their school ranked in the top half of all independent schools. They had already learned to read and write in both French and English a wide variety of genres, including daily journaling activities, book reports, written poems and much more. The class was held for one hour on Thursday afternoons after regular school classes were out in the school computer lab in the basement.

The Moliere children, on the basis of their older age, longer and more extensive training in genre and composition forms often taught in schools, would be expected to make more sophisticated spriting products than both groups of relatively younger Umoja children. Differences between them cannot be attributed to their interaction with spriting. However, when certain behaviors, processes or products emerge in common between the schools—even against the predicted differences, these commonalities should be subjected to further analysis and interpreted in light of what spriting offers. In sum, how might spriting allow each child to exercise their current literacy skills, develop new abilities, and challenge themselves in powerful ways?

By acting as researcher, teacher, and technologist, I was able to exercise great flexibility in focusing and refocusing literacy activities on the fly, the technology on a weekly basis, as well as the classroom activities and forms of student participation with each other and with the computer. I judged an activity successful when the children had energy and long-term focus, spent a lot of time on editing, or requested more time. The definition of acceptable spriting work was something that the students and I continuously negotiated throughout the thirteen weeks at both schools as what could both sustain their interest and mine. However, the most unexpected results emerge from observing children's great enthusiasm for activities I did not suggest and was initially confused by. For example, although I initially framed a talkument as an intermediate product on its way to becoming a text, the children embraced talkuments as final products. Their energy and interest in making talkuments in turn encouraged me to adapt the technology to produce a better talkument product and to involve writing in ways that supported their spriting process rather than the other way around.

Data collection included video records of all spriting activities, saved daily versions of spriting work, saved records of interactions with the technology, examples of the students' writing from their regular classroom teachers, and detailed field notes. In total, the children from both schools produced a total of 197 compositions. To narrow and balance the descriptive analysis, I chose three to four solo compositions to transcribe from each child, a total of 67 compositions. Selection was based on the work's resemblance to or distinctiveness from written genres; its virtuosity relative to the child's other compositions; and the importance the child ascribed to it. The latter two factors often led to the same selection, as the work that interested them most was also the one that they worked hardest at. These 67 talkuments and all the recordings made in the process - even if ultimately deleted - were transcribed. Transcription includes notation of (filled) pauses and breathes, false starts, incomplete and unusually pronounced words using embedded XML marks, used for automated processing purposes, as well as to remind the reader that this is an imperfect and partial translation of speech, impossible to capture fully in all of its *material repleteness*. When children are trying the hardest to transform thought, feeling and intention into words is when their speech is most fragmented (Chi, 2000). The transcription attempts to draw attention to these periods of intense linguistic construction.

Since the time when I performed this work, podcasts, moblogging and other voice publication methods have emerged as legitimate and authentic purposes for real-time voice composing, making this work all the more relevant as the means of distribution of such work are emerging. Understanding the importance that children attribute to spriting and talkuments becomes even more critical for how it might inform current literacy practice and pedagogy, and how it might presage future literacy trends.

## Overview of the Composition and Editing Actions

Looking across the 67 transcribed talkuments including both individual and collaboratively composed talkuments, there is strong evidence that all children, ages 5-10, composed at length, reviewed and edited their work. Considering that at the time the youngest children were working hard to write their names, this result alone, when compared to standard letteracy development comparisons across the same ages, is a striking result. Within this general trend, there are indications of developmental differences. In order to see these developmental trends, the data is organized in to three different groups descending in average age: Molière children (ages 8-9), the older Umoja children (ages 6-8, 10), and the younger Umoja children (ages 5-6).

**Table 1 Record and Play composing moves, Recording lengths, and Total Words.**

| | | Record Actions (#) | Play Actions (#) | Record then Play Consecu. | Recording Len. Each (sec) | Talkument Len. Total (sec.) | Words Total |
|---|---|---|---|---|---|---|---|
| **Moliere** | N | 31 | 31 | 30 | 31 | 32 | 32 |
| | AVE | **18.29** | **31** | **8.7** | **37.2** | **349.3** | **359.2** |
| | SD | 17.7 | 30.3 | 7.2 | 34.4 | 216.6 | 276.4 |
| **Umoja older** | N | 25 | 22 | 21 | 25 | 25 | 24 |
| | AVE | **8.6** | **17.2** | **4.5** | **46** | **273** | **222.3** |
| | SD | 6.1 | 14.3 | 4 | 47.8 | 214 | 168.2 |
| **Umoja younger** | N | 10 | 10 | 10 | 10 | 10 | 10 |
| | AVE | **9.8** | **30.3** | **6.2** | **44.4** | **358.9** | **256.2** |
| | SD | 6.3 | 25.6 | 4.6 | 30.5 | 268.1 | 207.2 |

In Table 1, the average number of Record and Play actions per talkument are presented for each group. The Molière children, as predicted by their older average age, made an average of 18 recordings per talkument, and listened to their spriting on average 31 times per talkument. The Umoja children made more or less the same number of recordings per talkument.

While the Molière children tended to make more recordings, they also tended to make slightly shorter recordings (37 sec on average in Table 1 column "Recording Len. Each"), tending towards a more bricolage, piecemeal construction process than all the younger Umoja children. This "bits and pieces form a whole" style emerged after a few weeks in a couple of the Moliere children who desired extreme control over their talkument product. Most children at both schools tended towards long spoken expositions, with a few gross post hoc edits.

As presented in Table 1, total talkuments lengths ranged from 270 to 360 seconds for all children—with the youngest children producing the longest talkuments! This is exactly the opposite of what one would predict in writing, and an extremely important insight into what spriting can offer literacy development – composition length is no longer a bellwether indicator of literacy development. Length of talkument, however, does not predict the number of words used. Older children had more fluency with spoken language and greater grasp of compositional planning, thus averaging a little more than one word per second (359 words in 349 seconds) while the youngest averaged approximately 2/3 a word per second (256 words in 358 seconds).

Older children tend to have composition plans that exceed a single recording action. In Table 1, the 'Record then Play' column indicates the sequential relationship between recording and playing actions. The Molière children immediately listened to only 8.7 out of the average 18.29 recordings they made, while the younger Umoja children listened immediately to 6.2 out of an average 9.8 recordings—2/3 of all their recordings. I observed the older children make several recordings consecutively and then listen to a group of them, while the younger children composed in a record-listen-record-listen cycle, the next move inspired by what they have just heard. Interestingly, the youngest children also listen more often to their own recordings—in Table 1 their record-play ratio is 1:3 compared to the older children's 1:2. Thus, there are indications of a developmental trend towards making shorter, more numerous recordings that are listened together as composites (not individual recordings), in a review cycle that occurs less frequently.

**Table 2 Average editing moves per talkument.**

| | | *Delete* | *Move* | *Split* |
|---|---|---|---|---|
| **Moliere** | N | 30 | 17 | 14 |
| | AVE | **11.8** | **8.8** | **9.1** |
| | SD | 11 | 12.6 | 20.6 |
| **Umoja older** | N | 20 | 4 | 10 |
| | AVE | **8** | **4.8** | **3.2** |
| | SD | 6.6 | 4.9 | 3.2 |
| **Umoja younger** | N | 9 | 5 | 7 |
| | AVE | **9.9** | **3** | **5.1** |
| | SD | 10.5 | 2 | 4.7 |

Children engage in editing moves through spriting even before they know how to write. As they get older, they engage in even more editing moves. Table 2 shows that the younger Umoja children delete and move content, and make splits. The Molière children make on average 12 deletions and move sausages nearly 9 times *per talkument*, and they make finer edits by splitting content when necessary. The standard deviation for split edits is very high because children used splits in only certain talkuments and certain situations. The older Umoja children fit somewhere in the middle of these two groups as they begin to compose lengthy and purposeful talkuments. Spriting might be treated now as a complementary process to writing in the elementary years, both spriting and writing processes eventually to equalize in length and complexity.

## Manifestations of Planning Processes in Spriting

The spriting process affords new insights into composition generally, and might provide new insights into the general theory of composition process—heretofore examined only through the window of writing behavior. For example, planning processes in spriting are very different from what children seem to experience when writing. In spriting, several children learned that planning is a time for *pausing*.

I encouraged the children to take intentional moments for thinking. This kind of thinking activity is often called "planning" in the writing process literature, but has unfortunately been co-opted in most school settings with highly structured activities for purposes of global planning (e.g. outlines, templates, concept maps, et cetera) that should occur "first" in the composition process. Though these devices can be useful, they do not supplant the need for thinking throughout the composition process, sometimes called "local planning" in writing composition research. Ironically, the value of these kinds of global planning devices might be more recognizable, and even more necessary, in the spriting process than in the writing process, when children discover for themselves how useful broad composition plans are when writing mechanics are not the primary obstacles.

Many children become frustrated when spriting—and even frightened away from trying to sprite—by the immediacy of the task. They find that the words they expect to be there are *not* there. They fall into a silence they find uncomfortable, even unbearable. Because children talk easily when they respond to adult questions or when they converse with their friends, they seem to expect that words will always emerge to accomplish the task they have set for themselves. When they watch television, they see people who always know what to say. They see politicians answer questions no pause. We have few public models of people pausing to think before talking. Therefore, it is not surprising that children, even if they have very little experience writing or spriting literate compositions, do not expect that thinking could take time.

In writing, the time to get something down on paper or screen is much longer than speech (300% to 400% longer even for an expert typist). Therefore, the time spent thinking and planning comes for 'free' while a child struggles with mechanical production. I saw many children struggle initially with the fact that they could not say everything they wanted to in one long, synchronous recording, as they felt they did when writing. They would get frustrated at their self-perceived lack of fluency, erase everything and start over again. The idea that thinking is doing something had not seemed to occur to them before. Writing may be less conducive to cultivating the habit of mind to pause to think during composition. Spriting seemed to help students learn aspects of composition that they may have performed implicitly while writing, due to the slower mechanics, but had not realized they did so in such a way as to use this cognitive skill flexibly.

In spontaneous speech, Frieda Goldman-Eisler (1958) claims that pauses provide an external window on internal cognitive processes, and that pause lengths increase with task abstraction and 'explicitness' – a feature correlated with how familiar/unfamiliar the conversants are with each other. When I began the design research, I believed that pauses, filled (e.g. um, uh) or empty, should be reduced and minimized respectively as a feature of editing a talkument from intermediate to final product. I learned from the children that pauses – both filled with a delightful variety of sounds and empty – could be indispensably important to a talkument. Next, I tell two learning stories of how pauses were important in spriting. Neither describes the general case of learning to pause but instead complicates the notion of what pausing to think should be. These stories demonstrate that a single, global approach to pauses in spriting (pedagogically, technological support, etc.) would not only be difficult to achieve in practice, but might also be undesirable.

## Madeline Likes Her Pauses

In the fourth week of spriting class at the Molière Elementary, I devised an activity intended to focus the children on editing, and in particular to remove what I considered long and unnecessary pauses. They were to create a talkument with three short stories. If any story exceeded four spriting lines in length, they were to edit it down. They could consult with friends to figure out what might be eliminated. The best story was to be placed first in the composition (by dragging and dropping). Although complicated, the children understood these directions and produced some material that was amongst their best work for the entire spriting class.

Eight-year-old girls, Madeline and Emily, finished quickly and were the first to listen to each other's compositions. I instructed them also to point out long silences to their partner. Emily must have commented to Madeline that her composition had too many long silences. Madeline motioned me over to discuss this.

My composition has many long silences. Do I have to remove them? She said. Actually, I like the silences.

Why is that? I asked.

When I talk I often use long silences.

Well, just because your talk has long silences doesn't mean you want your composition to be exactly the same. The person listening to this will not be able to look at your face when hearing your talkument, they can only hear your voice. Sometimes when you aren't saying anything, your face says something instead. But in a talkument you can't use your face. So make sure that the silences are useful.

Clearly I didn't get her point, so she tries to restate it differently.

I also use a long pause before something scary. So, I prefer to keep them.

It hits me. She's referring to the use of pauses for dramatic effect, a performance issue an oral storyteller would be very concerned with. Using pauses for dramatic effect demonstrates a mature sense of verbal storytelling. Pauses are not always emblematic in spriting of an internal cognitive struggle to produce words. If these dramatic pauses were removed—or even shortened—automatically by the spriting technology, Madeline's mature compositional purposes would be destroyed.

Due to this and other experiences that demonstrated a usefulness and meaning to pauses that I had not considered before, I disabled the SpriterWriter's technological support for automatically shortening long pauses.

## Niesha's Beats as Pause

Niesha, an eight-year-old child at Umoja Elementary, had a gift for beats. These beats consisted of lip smacks, purses, sucks, kisses, clicks, hummed bars from songs, and more, all merged together into a drum rhythm that accompanied her spriting. Many of the Umoja children were able to create inventive, spontaneous rhythms with their voices, whereas none of the Moliere children could. Certainly a familiarity with rhythmic talk as heard in rap and hip-hop contribute to their able improvisation. But Niesha was particularly gifted amongst her classmates. She enjoyed making her compositions a continuous, unbroken weave of rhythm, talk and song. For example, here is a transcribed excerpt from one of her compositions:

> I like (.) hmm {VOC mouth sounds} (7.4) I like spaghetti, {VOC mouth sounds} I like chicken {VOC mouth sounds} spicy chicken {VOC mouth sounds} I like Chinese food, the rice the Chinese rice and the (.) red - no the orange chicken (inaudible) they are so good if you eat them your mouth bam! if you eat the orange chickens. {VOC mouth sounds}

This example is the first in a series of talkuments in which she uses descriptive and journal-like talk coupled with these drum-like mouth sounds and excerpts from songs to describe foods she likes and its effect upon her. In between each mention of a food, she made what she called a 'beat.' Beats for Niesha include sounds she makes with her mouth, song excerpts, and small bits of Americana like "Peace", "Ho Ho Ho Merry

Christmas". Her beat is nearly impossible to transcribe to text. The (inaudible) mark in particular, and the {VOC mouth sounds}, often stands for sections of spriting that simply have no textual equivalent.

One of Neisha's primary goals was to make long talkuments; she bragged about the great length of her spriting multiple times to me and to her teacher. How ridiculous our goals for mature writing, such as length, appear when they are transplanted into another media like spriting in which they are so easily achieved! Her composition approach was typical of most children at both schools. She first recorded one or two continuous recordings and then edited these by adding or deleting a few major narrative chunks. I dubbed this editing approach "chunk insertions". This composition strategy required great fluency, and children had to devise ways to plan what to say next while they were recording. Niesha appears to use beats as pausing time. Her nearly automatic ability to perform beats allowed her to time to construct her next thoughts. The effect of this strategy is a kind of stream of consciousness flood of verbal and sound images from Niesha's life experience.

## Discussion

The quantitative descriptive overview and qualitative examples are meant to suggest very different kinds of conclusions than those that have dominated literacy research, particularly with respect to the ascribed importance of writing and text in literacy development. When children are learning to speak, they explore issues both great and small simultaneously. But only a few years later when they begin school, they focus intensely on the letterate issues – the smallest units of writing composition, and neglect practice of literate concepts for several more years. The introduction of spriting at beginning levels of letterate and literate development appears to allow children practice with extended forms of literacy and habits of mind such as review and editing well before they can properly execute the full character set.

One specific issue discussed here is the importance of learning a literate habit of mind – that is, to recognize the need to make composition plans. In written discourse, planning comes for free with the longer written mechanic requirements. But spriting speeds up the composition process to a pace that resembles spontaneous speech, but lacks most if not all of the social supports. To characterize the importance of pauses in literacy development, it would be helpful to distinguish two different kinds of learning. The first is actually learning to do something; for example, learning how to sing. The second kind of learning is to be conscious of *how* one actually does that thing. This distinction might be related to Karmiloff-Smith's concepts of implicit and explicit knowledge (1992), sometimes also referred to as meta-discourse or meta-awareness. She writes that the ability to do something (implicit knowledge) can disappear for a time as the knowledge of how to do this thing shifts to making hypotheses about how one actually does this (explicit forms). Although both types of learning are not required for action (obviously, just learning how to do something is often enough), it is sometimes critical to know how one actually does it in order to deploy the ability when the usual supports are not available. Spriting allows students to discover habits of the literate mind that they may already do when writing, but do not understand or identify as actions involved in composing. Spriting forces them to think about—and potentially develop new hypothesis for— how they go about the process of composing. If children have the ability to make more aspects of the composition process explicit, they can control these skills in both their written and spoken compositions.

## References


Bereiter, C., & Scardamalia, M. (1987). An Attainable Version of High Literacy - Approaches to Teaching Higher-Order Skills in Reading and Writing. *Curriculum Inquiry, 17*(1), 9-30.

Brown, A. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *The Journal of the Learning Sciences, 2*(2), 141-178.

Chi, M. (2000). Self-Explaining Expository Texts: The Dual Processes of Generating Inferences and Repairing Mental Models. In *Advances in Instructional Psychology* (pp. 161-238). Mahwah: Lawrence Erlbaum.

Collins, A. (1992). Toward a Design Science of Education. In E. Scanlon & T. O'Shea (Eds.), *New Directions in Educational Technology*. New York: Springer-Verlag.

diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sciences, 13*(1), 77-103.

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences, 11*(1), 105-121.

Goldman Eisler, F. (1958). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Haas, C. (1989). How the Writing Medium Shapes the Writing Process - Effects of Word-Processing on Planning. *Research in the Teaching of English, 23*(2), 181-207.

Jones, I. (1998). The Effect of Computer-Generated Spoken Feedback on Kindergarten Students' Written Narratives. *Journal of Computing in Childhood Education, 9*(1), 43-56.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, MA: MIT Press.

Olson, D. R., & Torrance, N. (1985). *Language, Literacy and Mental States*. Canada; Ontario: Ontario Inst for Studies in Education Toronto.
Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. NY: Basic Books.
Purcell-Gates, V. (1991). Ability of well-read-to children to decontextualize/recontextualize experience into a written narrative register. *Language and Education: An International Journal, 5*, 177-188.
Shankar, T. R. (2006) "Speaking on the Record: A Theory of Composition." *Computers and Composition.* In C. Ball and B. Hawk (Eds.) Sound in / as Compositional Space: A Next Step in Multiliteracies.
Shankar, T. R. (2005). *Speaking on the Record.* PhD, Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

## Acknowledgments

My thanks to the three anonymous reviewers for their helpful critiques, and my gratitude to my former advisor, Seymour Papert, who legitimized my dream and made this work possible.

# Using Knowledge Space Theory to Analyze Concept Maps

Laura A. Cathcart, Mike Stieff, Gili Marbach-Ad, Ann C. Smith, and Kenneth A. Frauwirth
University of Maryland, College Park, Maryland
cathcart@umd.edu, mstieff@umd.edu, gilim@umd.edu, asmith@umd.edu, kfrauwir@umd.edu

**Abstract:** This study examines use of knowledge space theory as a novel method of analyzing concept maps. Concept mapping is a technique for expressing relationships between ideas, using two-dimensional node-link diagrams to visually display relationships between ideas. We introduced concept mapping as a voluntary exercise in an upper-level undergraduate immunology course. The students were assigned ten concept maps (for which concept lists were provided) at intervals during the semester-long course. We utilized knowledge space theory (Folmagne & Doignon, 1988) to systematically analyze and compare concept maps drawn by students to their instructor's "expert" concept maps. Using this novel analysis method, we were able to reveal students' level of understanding of course material, changes in student knowledge across time, as well as identify students' alternative conceptions. We found knowledge space theory a productive tool for systematically analyzing and comparing students' concept maps across time and to expert's concept maps.

## Introduction

Concept mapping is a technique for externally depicting students' internal knowledge structures. Students construct concept maps by using labeled linking lines (propositions) to show relationships between ideas or objects (concepts). The process of constructing a concept map helps students build a deeper and more meaningful understanding of content material (Novak & Gowin, 1984). Concept maps are a useful pedagogical tool because they allow teachers to visualize students' knowledge structures. This allows teachers to structure lessons to expand upon students' existing knowledge structures (Edmondson, 2000; Marbach-Ad et al., 2007; Okebukola, 1990). Despite the successful instructional use of concept maps, they have proven difficult to use for assessing student knowledge.

In part this difficulty is due to the various concept map assignments, instructions and analytical techniques or scoring procedures. Some concept mapping assignments ask students to generate the important concepts for a topic, others ask students to fill in the propositional statements on the linking lines on a pre-drawn concept map, and still others ask students to construct a map from a list of important concepts. Concept mapping instructions might ask students to draw hierarchical relationships and crosslinks between concepts whereas others simply ask students to connect concepts together using propositions. Some concept map analytical techniques evaluate maps based on the degree of hierarchy and crosslinking between hierarchies, other scoring procedures are based upon the number of concepts and propositions in the map, and still other methods attempt to rate the quality of the propositions and crosslinks in the map. Ruiz-Primo and Shavelson (1996) wrote a review highlighting the myriad array of concept mapping tasks and analytical techniques or scoring procedures. In their paper, Ruiz-Primo and Shavelson call for researchers to more carefully design concept mapping assessment tasks to align with cognitive theories. They also stress the need to establish the validity and reliability of analytical techniques for concept map assessments before widespread implementation of the assessment.

Many researchers (e.g., Okebukola, 1990; Markow & Lonning, 1998; Trowbridge & Wandersee; McClure et al., 1999) have experimented with different concept mapping tasks and analytical techniques to develop concept mapping assignments that provide effective assessments of student knowledge. Novak and Gowin (1984) outlined a quantitative scoring procedure for hierarchical concept maps in their original book introducing the concept mapping technique. In this scoring method, each valid proposition was assigned 1 point, each hierarchical level of the map was assigned 5 points, and each valid and significant crosslink between branches was assigned 10 points. The sum of these points established the overall score for the concept map. Novak and Gowin's scoring technique is useful because it differentiates between simple and complex maps numerically; however, the technique depends upon the hierarchical nature of the concept map. In contrast, Kinchin, Hay, and Adams (2000) developed a qualitative method of analyzing student's concept maps. The method involves determining if the concept map's structure is best described as a spoke, chain, or net of propositions connecting concepts. This technique allows for a rapid determination of the complexity of the concept map. However, it does not account for the quality of the concept map. McClure et al. (1999) conducted a study in which they attempted to establish the reliability and validity of a relational scoring procedure for concept maps. The technique involves assigning a numerical quality rating to each proposition in a concept map.

Similar to Novak and Gowin, the sum of the numerical quality ratings for each proposition was used to represent the score for the concept map. This technique is useful because it incorporates the quality of each proposition into the concept map score. However, the relational scoring technique only provides information about a student's knowledge relative to other students. The relational scoring technique does not provide instructors with information about students' alternative conceptions or concepts that may need further instruction.

Despite the utility of each of the above analytical techniques, many questions remain about the theoretical rationale behind such techniques and the validity and reliability of each procedure. There is a need for the development of alternative procedures for systematically analyzing concept maps to reveal information about students' knowledge structures. Knowledge space theory (Folmagne & Doignon, 1988) offers one potentially useful procedure for systematically analyzing concept maps. Knowledge space theory (KST) is a mathematical theory that allows researchers to represent the state of a student's knowledge, referred to as a knowledge space, at a precise point in time using a series of assessments. The first step in determining a student's knowledge space involves carefully designing a set of questions that each measure knowledge of a discrete concept. The set of questions is administered to the student and scored for correctness. The unique combination of concepts, represented by the correctly answered questions, comprises a students' knowledge space. KST is useful to educational researchers because it allows for the systematic comparison of students' knowledge in a particular domain across time and compared to experts (Folmagne & Doignon, 1988). KST was originally designed to help researchers examine students' knowledge spaces in mathematics domains, but has also been shown to successfully represent student knowledge in science (Taagepera et al., 1997). By administering the same survey throughout a semester of chemistry, the researchers determined potential learning pathways. That is, students appeared to learn additional concepts in discrete series. Figure 1 illustrates an example critical learning pathway for chemistry students.

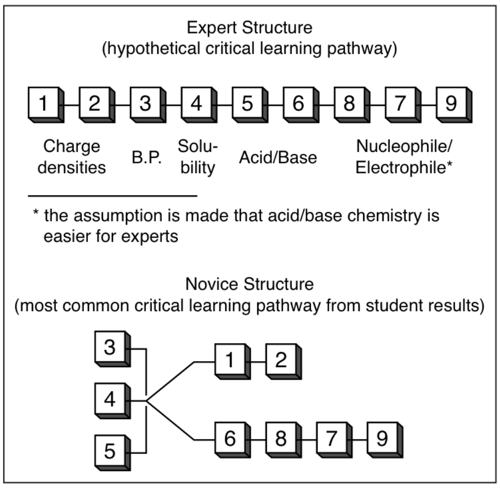



Figure 1. Critical learning pathways for novices and experts. (Taagepera & Noori, 2000).

In the present study, we applied a modified KST framework to analyze concept maps created by students enrolled in an undergraduate immunology course. We focused on examining the composite of students' knowledge spaces across time and to the instructor's knowledge spaces. In contrast to the KST approach of identifying correct answers to assessment questions, we built student knowledge spaces by identifying the presence or absence of propositions linking concept pairs on concept maps. Due to the voluntary nature of the concept mapping assignment, our data was not conducive to comparing change in individual students across time. Rather, our results indicate that analysis of concept maps with the KST framework permits the identification of the relative frequency of domain concepts in a population of students as well as notable alternative conceptions. The KST framework is also useful for highlighting similarities and differences between instructor's and students' concept maps.

## Methods

An instructor employed concept maps as instructional aids in a team-taught undergraduate introductory immunology course. The course consisted of two 1-hour lectures and one 1-hour discussion per week. Students were given a brief overview of the concept mapping technique and the instructor demonstrated construction of a sample concept map at the start of the course. On 10 separate occasions in the course, students were assigned concept maps on specific lecture topics: B Cell Development, Germinal Centers, Major Histocompatability

Complex, T Cell Development, Costimulation, Cytokines, Autoimmunity / Tolerance, Alloreactivity / Immunosuppression, Vaccines, and Immune Responses to Tumors. For each map, the instructor provided students with 11 to 17 concepts (M = 14) central to the corresponding lecture topic to use while constructing each concept map. Note that several of these concepts were present on multiple concept map assignments. Students received 1% extra credit towards their final course grade for completing 5 of the concept maps. After students submitted their concept maps, the whole class received a concept map on the same topic created by the professor. We analyzed 208 concept maps collected from 89 students enrolled in the course of 98 students. Of the 89 participants, 52 completed one or more concept maps and 37 completed five or more concept maps. Table 1 illustrates the total number of concept maps completed for each lecture topic. Due to the voluntary nature of the assignment, a different set of students completed concept maps for each lecture topic.

Table 1. Number of submitted concept maps for each lecture topic.

| **Lecture Number** | **Lecture Topic** | **Number of Submitted Concept Maps** |
|---|---|---|
| 6 | B Cell Development | 27 |
| 6 | Germinal Centers | 25 |
| 8 | Major Histocompatability Complex | 24 |
| 11 | T Cell Development | 23 |
| 13 | Costimulation | 17 |
| 15 | Cytokines | 18 |
| 21 | Autoimmunity / Tolerance | 20 |
| 22 | Alloreactivity / Immunosuppression | 27 |
| 24 | Vaccines | 17 |
| 25 | Immune Responses to Tumors | 10 |
| Total = 26 | | Total = 208 |

The concept maps constructed by the instructor and students were exclusively network concept maps that included “a semantic network with concept nodes linked directionally by labeled lines (arrows) to produce propositions” (Ruiz-Primo & Shavelson, 1996, p. 572). Unlike hierarchical concept maps, network concept maps are not arranged in a hierarchical nature (Novak & Gowin, 1984). Figure 2 illustrates an example student concept map for the B Cell Development lecture topic.

All of the concept maps were analyzed using the KST framework. We used the propositions in each concept map in place of the discrete conceptual questions used in traditional applications of KST. For each set of concept maps from a lecture topic, every unique proposition was assigned a letter or number code. All propositions generated by the instructor (those that appeared on the instructor’s map for the topic) were assigned a number; the list of numbered propositions for each topic was set by the instructor’s concept map. This list was then compared to each student’s concept map to identify the instructor-generated propositions on students’ concept maps. All propositions generated by the students (those that appeared only on students’ maps for the topic) were assigned a letter. The list of lettered propositions for each topic grew as each student’s concept map was analyzed. This list was then compared to each student’s concept map to identify the student-generated propositions. Akin to the techniques used in traditional KST, the unique combination of propositions, represented by numbered and lettered codes on each concept map, represent a student’s knowledge space for that lecture topic.

In this study, we did not focus on comparing individual student’s knowledge spaces; instead we focused on comparing a composite of students’ knowledge spaces across time and to the instructor’s knowledge spaces. The frequency of numbered (instructor-generated) and lettered (student-generated) propositions was calculated for each lecture topic. We constructed composite maps (See Figures 3 and 4) for each lecture topic to help interpret and visualize the changes across time and the differences between the students’ and the instructor’s knowledge spaces. Instructor-generated propositions were highlighted in **bold** text on the composite map if they appeared on less than 50% of student concept maps. The threshold was set at 50% indicating the potential need for instructional intervention in the classroom as opposed to individual student errors. Correct student-generated propositions that were not indicated by the instructor, were highlighted in CAPITAL text. Incorrect student-generated propositions were highlighted in *ITALICIZED CAPITAL* text if they appeared on more than 25% of student concept maps. The threshold was set at 25% indicating the possible presence of a shared alternative conception among students as opposed to an individual student's misunderstanding.

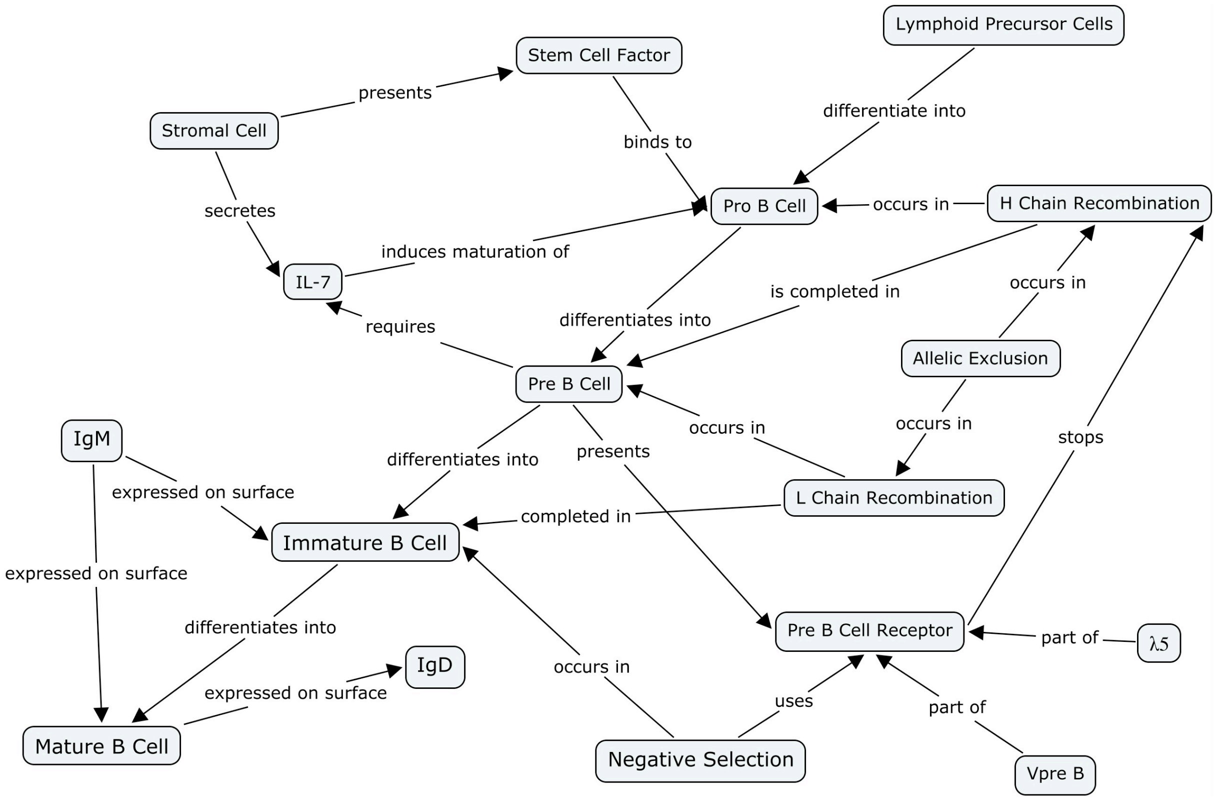



Figure 2. Example Student B Cell Development Concept Map. Concepts are the ideas enclosed in rounded boxes. Propositions are represented as directional arrows labeled with a linking phrase.

## Results

Knowledge space theory proved useful for comparing students' concept maps across time and to their instructor's "expert" concept maps. Generating the composite maps for each concept map lecture topic, systematically summarized the knowledge spaces represented by the students' concept maps in comparison to the instructor's "expert" concept map. By calculating the frequency with which the propositions from the instructor's concept map appeared on student's concept maps, the researchers were able to systematically highlight the differences in knowledge structures between the students and the instructor.

Despite the uniqueness of each concept mapping lecture topic, we were able to compare students' knowledge spaces across time, because some of the same concepts appeared in several of the assigned concept maps for different lecture topics. The B and T cell development concept maps provided a unique opportunity to compare the development of students' knowledge spaces over the course of the semester. The B cell development lecture topic concept map was the first map assigned during the semester. The T cell development lecture topic concept map was the fourth concept map assigned during the semester, after the students had taken and received feedback for an exam on B cell development. B cells and T cells both initially share similar development processes that diverge later in the developmental process. However, several of the processes that regulate the development of B cells are common to T cells as well. This unique situation allowed the researchers to analyze the change across time of students' knowledge spaces on the common processes in B and T cell development. Figures 3 and 4 illustrate the composite maps for the B and T cell development lecture topics. KST analysis of the B and T cell development concept maps revealed both positive and negative changes in students' knowledge and relatively stable student knowledge structures.

Comparison of the B and T cell development concept maps revealed an improvement in students' understanding of some concepts across time. Analysis of the B cell development composite map indicates that only 30% of the students included information about the interaction of the lymphoid precursor cells with the stromal cells. The T cell development composite map indicates that information about the interactions between the lymphoid precursors and the stromal cells appeared in 95% of the students' knowledge spaces. This dramatic change in the appearance of a proposition indicates a significant improvement in students' understanding across time.

KST was also useful for identifying positive changes in students' knowledge, such as student-generated propositions that were not emphasized by the instructor. More than a quarter of students' concept maps from the T cell development concept maps included student-generated propositions about the interactions between the lymphoid precursor cells and the stromal cell effector molecules, IL-7 and stem cell factor. These interactions were not indicated by a significant percentage (more than 25%) of students on the B cell development concept maps. Despite the accuracy of these student-generated propositions, the instructor had not included the propositions on his "expert" concept map because he was not trying to emphasize these particular relationships between the concepts. KST analysis proved useful for highlighting improvements in students' knowledge that would not necessarily have been identified using other assessment techniques.

Using KST to analyze students' concept maps also allowed researchers to show the change over time in the frequency of alternative conceptions in students' knowledge spaces. An average of 50% of students' B cell development knowledge spaces indicated a correct understanding of the role of allelic exclusion in B cell development. The appearance of the correct understanding of the role of allelic exclusion had decreased to less

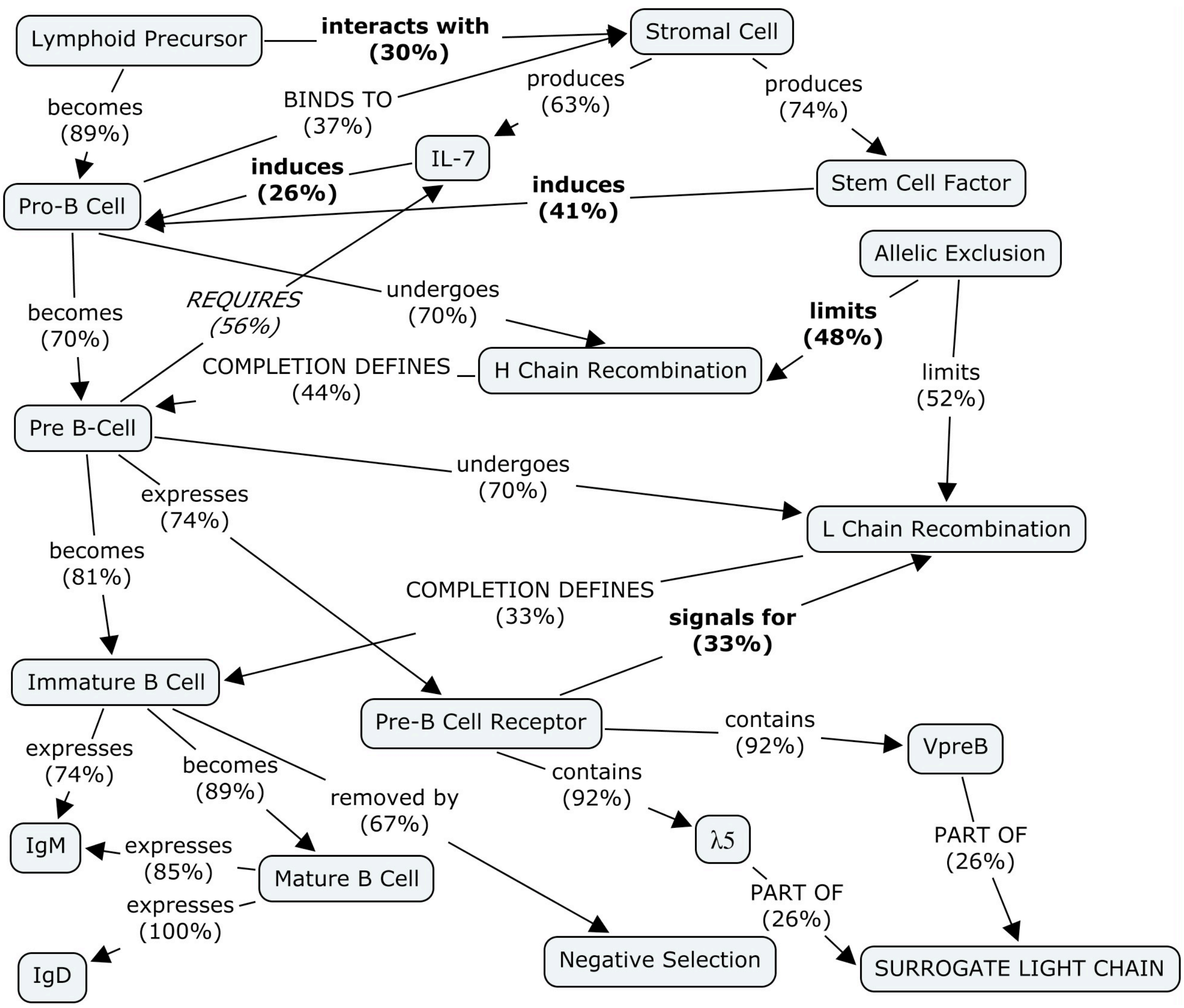



Figure 3. B Cell Development Composite Concept Map. Numbers in parentheses represent the percentage of student concept maps containing each proposition. Normal text represents concepts and propositions generated by instructor. **Bold** text represents propositions generated by the instructor that appear on less than 50% of student concept maps. CAPITAL text represents correct concepts or propositions generated by more than 25% of students. *ITALICIZED CAPITAL* text represents incorrect propositions generated by more than 25% of students.

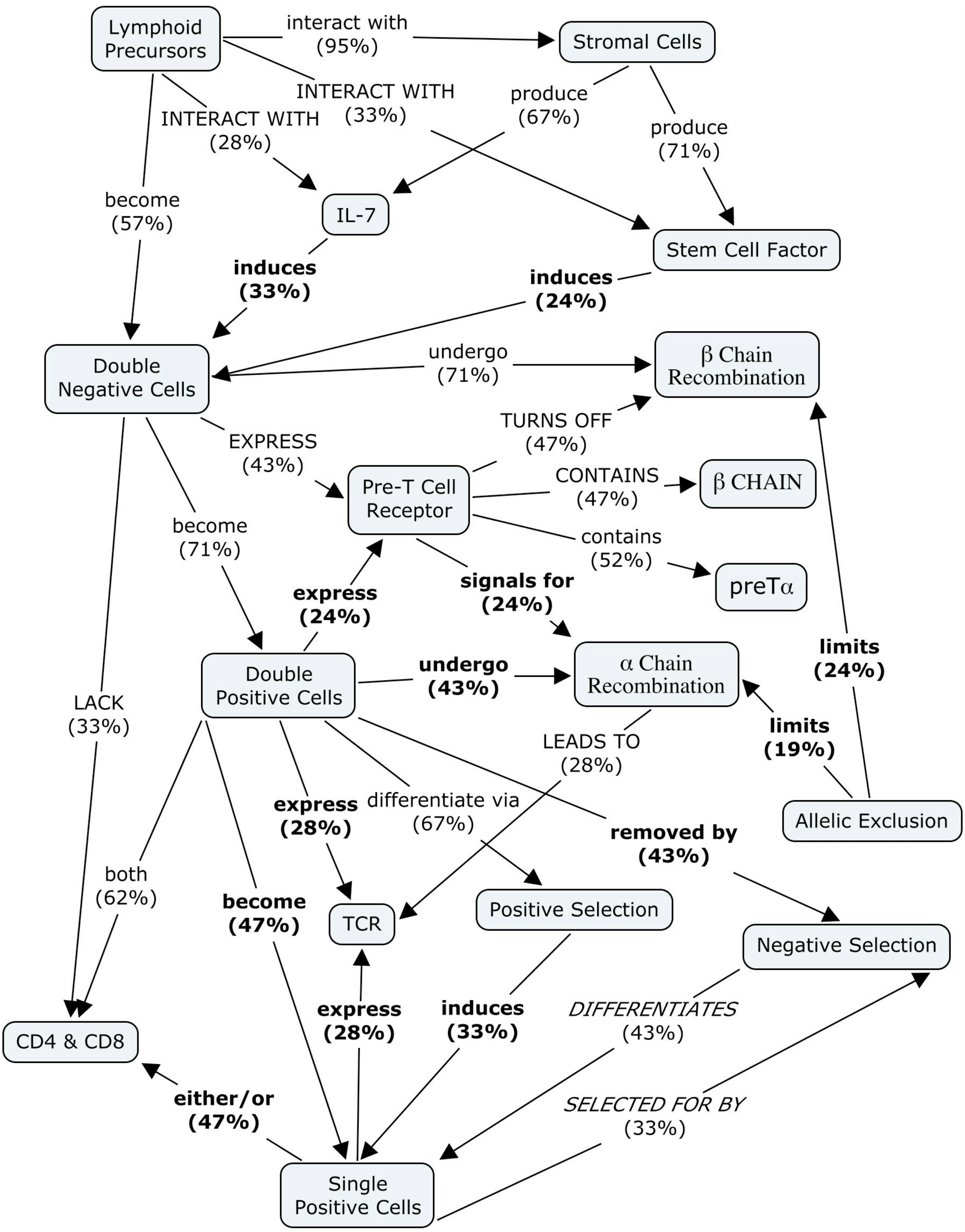



Figure 4. T Cell Development Composite Concept Map. Numbers in parentheses represent the percentage of student concept maps containing each proposition. Normal text represents concepts and propositions generated by instructor. **Bold** text represents propositions generated by the instructor that appear on less than 50% of student concept maps. CAPITAL text represents correct concepts or propositions generated by more than 25% of students. *ITALICIZED CAPITAL* text represents incorrect propositions generated by more than 25% of students.

than 25% in the students' T cell development knowledge spaces. A possible explanation for this change is that on average, students had a weak understanding of allelic exclusion in B cell development. T cell development is a more complicated process compared to B cell development and perhaps this exacerbated students' misunderstandings of allelic exclusion.

The technique of KST analysis of concept maps was limited by its ability to detect students' alternative conceptions at low frequencies. The process of negative selection was correctly indicated in less than 50% of students' T cell development knowledge spaces. Two incorrect alternative conceptions about the role of negative selection appeared in more than 25% of students' T cell development knowledge spaces. Our initial analysis did not reveal alternative conceptions about negative selection in students' B cell development knowledge spaces. Upon further analysis, 10 different alternative conceptions about negative selection were identified in student' B cell development knowledge spaces; however, all appeared at frequencies less than 25%. The thresholds set by the researchers limit KST analysis of concept maps, but these thresholds are flexible and can be adjusted as needed to detect different frequencies of students' alternative conceptions.

KST analysis of the B and T cell development concept maps also revealed concepts for which students' understanding did not improve over time. There was no significant change in the appearance of propositions about the interactions between the two stromal cell effector molecules, IL-7 and stem cell factor, with the lymphoid precursor's respective progenitor cells, the pro-B cells or double negative cells. On the B cell development composite concept map, only 26% (IL-7) and 41% (stem cell factor) of students' maps included propositions relating the stromal cell effector molecules to the pro-B cells. On the T cell development composite concept map, only 33% (IL-7) and 24% (stem cell factor) of students' maps included propositions relating the stromal cell effector molecules to the double negative cells. KST analysis revealed problematic concepts for which student knowledge remained relatively stable and did not improve over time with further instruction.

By analyzing students' concept maps with KST we were able to extract very fine detailed information about changes over time in students' knowledge spaces for lymphocyte development. KST analysis highlighted both targeted and unexpected improvements in students' knowledge across time. Analysis of students concept maps by KST revealed alternative conceptions and allowed us to track the changes in frequency of these alternative conceptions as further instruction was given about the concepts over time. KST analysis also highlighted problematic concepts for which students' knowledge was relatively resistant to instruction and remained constant over time.

The above examples have strong technical content that can obscure the critical viewpoints generated by this research. The research above used examples of immunology based concepts; however, we propose that the KST method of analyzing concepts maps can be expanded to any subject at the primary, secondary, undergraduate and even graduate levels. Using KST to analyze concept maps enhances the instructional usefulness of concept maps for both students and instructors. KST is a powerful technique for analyzing concept maps to understand a student's knowledge space at a given time, elucidate misconceptions held by a significant proportion of the students, and identify correct student-generated ideas. KST combined with concept mapping assignments has the ability to improve instructor clarity and students' perception, increase students' understanding of concepts, and assist instructors in constructively addressing students' misconceptions.

## Conclusion

Concept mapping is an effective pedagogical technique; however, concept mapping has proven challenging to use as an assessment. The KST framework is a theory driven, systematic method of analysis that can be used to evaluate the content of concept maps. We found that analysis of concept maps with the KST framework is a productive method for examining change across time in students' knowledge spaces. We also found that analysis of concept maps with the KST framework effectively highlights details of the similarities and differences between students' and instructor's knowledge spaces.

KST analysis of concept maps is productive because it is sensitive to both the omission of conceptual ideas from students' knowledge spaces and the addition of alternative conceptions to students' knowledge spaces. The technique also allows instructors to rapidly and systematically identify major themes in students' knowledge from the vast array of information contained in students' concept maps. The range of analysis times for each set of concept maps from a lecture topic was between 1 and 4 hours. The analysis time depended on the number of concepts assigned for each lecture topic and the overall similarity between the students' and instructor's concept maps. The median analysis time for the sets of concept maps was approximately 2 hours. There is the potential for computer software to assist in the analysis of concept maps using KST and shorten the analysis time. An ideal program would generate a list of all propositions linking two concepts. From this list, the researcher could easily count the number of propositions similar to the instructor's and identify frequently occurring alternative conceptions.

One limitation of using KST to analyze students' concept maps is that the researcher sets the thresholds for highlighting differences between students' and instructor's concept maps. If the threshold is set too high, it can reduce the sensitivity of the technique to subtle differences in students' knowledge spaces. The analysis time of the KST method increases as the thresholds are lowered. Lowering the thresholds also dilutes the major findings of the analysis by highlighting minor, possibly less important differences in students' knowledge structures. Despite this limitation, KST enhances the pedagogical usefulness of the concept mapping technique as a formative assessment because it provides instructors with a productive framework for systematically analyzing students' patterns of understanding and misunderstanding.

Future research can expand upon our use of KST for analyzing concept maps to include comparison of individual student's concept maps across time to identify changes in individual student's knowledge structures. In addition, more research is needed to establish the reliability and viability of using KST for analysis of concept maps from a wide variety of subject matters across a range of levels of students. KST analysis of concept maps should be compared to other established techniques of measuring students' knowledge structures, such as traditional KST applications and other standardized assessment techniques. We hope that other researchers will elaborate upon and expand the application of the KST framework for analyzing concept maps to many different disciplines of study.

## References

Edmondson, K. M. (2000). Assessing science understanding through concept maps. In J.J. Mintzes, J.H. Wandersee, & J.D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 15-40). San Diego, CA: Academic Press.

Falmagne, J., & Doignon, J. (1988). A Markovian procedure for assessing the state of a system. *Journal of Mathematical Psychology, 32*, 232-258.

Kinchin, I. M., Hay, D. B., & Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, *42*(1), 43-57.

Marbach-Ad, G., Yarden, H., & Gershoni, J. M. (2007). Using the concept map technique as diagnostic and instructional tool in introductory cell biology to college freshmen. *Journal of Student Centered Learning, 3*(3), 163-177.

Markow, P. G., & Lonning, R. A. (1998). Usefulness of concept maps in college chemistry laboratories: students' perceptions and effects on achievement. *Journal of Research in Science Teaching*, *35*(9), 1015-1029.

McClure, J., Sonak, B., & Suen, H. (1999). Concept map assessment of classroom learning: reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, *36*(4), 475-492.

Novak, J.D., & Gowin, D.B (1984). *Learning how to learn.* New York: Cambridge University Press.

Okebukola, P. A. (1990). Attaining meaningful learning of concepts in genetics and ecology: An examination of the potency of the concept-mapping technique. *Journal of Research in Science Teaching*, *27*(5), 493-504.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33(6)*, 569-600.

Ruíz-Primo, M. (2000). On the use of concept maps as an assessment tool in science: What we have learned so far. *Revista Electrónica de Investigación Educativa*, *2* (1). Retrieved March 10, 2009 from: http://redie.uabc.mx/vol2no1/contents-ruizpri.html

Taagepera, M., Potter, F., Miller, G. E., & Lakshminarayan, K. (1997). Mapping students' thinking patterns by the use of the knowledge space theory. *International Journal of Science Education, 19(3)*, 283-302.

Taagepera, M., & Noori, S. (2000). Mapping students' thinking patterns in learning organic chemistry by the use of knowledge space theory. *Journal of Chemical Education*, *77*(9), 1224-1229.

Trowbridge, J. E., & Wandersee, J. H. (1996). How do graphics presented during college biology lessons affect students' learning? *Journal of College Science Teaching*, *26*(1), 54-57.

# Analyzing Equality of Participation in Collaborative Inquiry: Toward a Knowledge Community

**Abstract:** This paper shares preliminary findings from a designed-based study of the Knowledge Community and Inquiry (KCI) model for secondary science curriculum. We investigate the impact of the model on students' cooperative knowledge construction and their understanding of the science of climate change. Working closely with a science teacher, we co-designed a 10-week curriculum unit with small group and whole class collaborations across two participating classrooms. We present detailed analysis of the wiki content created by two groups of students, revealing a positive correlation between students' contribution to collaborative inquiry activities and their achievements in the curriculum. This finding suggests the need for increased scaffolding to support symmetric participation in collaborative inquiry so that both high and low achieving students benefit from such collaborative science curriculum activities.

## Purpose of the Study

Typical classroom instruction in science still focuses on rote learning of content, performance of problem solving, and coverage of standards and expectations (Slotta & Linn, 2009). Several traditions in the Learning Sciences, such as inquiry-based learning (e.g. Krajcik, Blumenfeld, Marx, & Soloway, 2000) and knowledge communities (Brown & Campione, 1994; Scardamalia & Bereiter, 2006), propose pedagogical approaches that would transform teaching and learning in science classrooms. Despite its success in increasing students' understanding of topics, inquiry-based learning still seeks strategies to promote collective knowledge construction within classrooms (Linn, Clark, & Slotta, 2003). By the same token, developing knowledge communities in classrooms has proven difficult due to lack of concrete guidelines for designing collaborative activities, as well as the challenges of bridging small-group collaboration with the wider community knowledge-base, and of sustaining long-term collaborative inquiry (Lehrer & Schauble, 2006).

This study investigates the effectiveness of a new hybrid model of collaborative inquiry situated within a knowledge community that enables students and teachers to engage in inquiry activities that are designed to deeply interconnect with a classroom-constructed knowledge base. To overcome the challenges of developing inquiry-oriented knowledge communities in secondary school science classrooms, the Knowledge Community and Inquiry model (KCI) proposes utilizing collaboration scripts (Kollar, Fischer, & Hesse, 2006) scaffolded by a technology-enhanced environment – in this case, a wiki. The KCI model builds on recent research that integrates structured collaborative patterns to complement collective knowledge construction and inquiry process to alleviate aforementioned challenges and foster inquiry-oriented knowledge communities (Hmelo-Silver & Barrows, 2008).

This research extends prior studies on KCI (Slotta & Peters, 2008) by integrating the model deeply into a grade 9 science curriculum unit concerned with climate change, and by adding a scripted collaborative component to the inquiry activities. The purpose of the larger study (now entering its second year of data collection) is to investigate the effect of the refined KCI model on students' cooperative knowledge construction and deep understanding of climate change concepts. The present paper will examine the quality of knowledge objects constructed by the students and the patterns of students' engagement in collaborative inquiry.

## Theoretical Framework

### Knowledge Communities

Learning science researchers advocate the notion of "Knowledge Community" to support learning. Of special interest to this research are the conceptual frameworks of Fostering Communities of Learners (FCL; Brown & Campione, 1994) and Knowledge Building Communities (KBC; Scardamalia & Bereiter, 2003) that have been intensively researched in the context of k-12 education. Although the two frameworks differ in their guiding principles, with FCL following a cycle of "research, share and perform" KBC pursuing idea improvement as a product of students' knowledge construction efforts, they can be synthesized along three dimensions.

Distributed cognitive responsibility: Traditional forms of instruction, such as lecture, are not conducive to helping students integrate knowledge for deep understanding of concepts (Brown, 1997). In a knowledge community approach, students become involved in articulating their own learning goals and planning the learning activities. Members of a knowledge community are not only responsible for the quality of their individual learning, but also are held accountable to communicate with other community members to scrutinize the quality of collective knowledge (Bielaczyc & Collins, 1999). Metacognition and reflection serve to increase students' awareness of how they learn and thereby enable self-reflection, critical inquiry, and flexible application of knowledge in novel situations.

Community knowledge-base: In a knowledge community approach, knowledge sources are no longer limited to the teacher and curriculum materials, but include knowledge artifacts identified or created by the students (Bielaczyc & Collins, 1999). A shared knowledge base promotes a diversity of ideas and facilitates access to those ideas in order to encourage the continued growth of community knowledge (Brown & Campione, 1994; Scardamalia, 2003). Structured collaborative activities intentionally distribute expertise among students, requiring them to negotiate their individual knowledge within collaborative groups and organize their new understandings in a common community knowledge base (Brown & Campione, 1994). Students' contributions to the community knowledge base will survive and be extended only if they are deemed important to the community (Scardamalia, 2003). Thus any artifact produced during such activities should carry the potential to be improved by others.

Pedagogical and technological scaffolds: To help teachers and students achieve the nuanced flow of ideas, activities and discourse within a knowledge community, researchers often introduce specific forms of guidance or scaffolds. These include collaboration scripts technology tools that guide students' knowledge building processes or discourse. In KB, for example, technology serves as a repository of ideas as well as a tool for collaborative knowledge construction (Scardamalia & Bereiter, 1999).

### Challenges for Fostering Knowledge Communities in Classrooms

Empirical research provides inconclusive results regarding the viability of the knowledge community approach in classrooms. The few published studies conducted on FCL-inspired classrooms report major challenges regarding instructional design and implementation (Sherin, Mendez, & Louis, 2004). Knowledge community frameworks propose epistemological more than procedural changes to learning and teaching. How will these principle translate into design guidelines for classroom practice? Sustained collaboration between researchers and teachers appears to be one enabling condition. Studies on knowledge building demonstrate the need for such partnerships to focus on principles to guide their design and implantation (Zhang, Scardamalia, Reeve & Messina, 2009). Despite some empirical evidence of effectiveness, knowledge community frameworks have not received widespread popularity among researchers and practitioners. A potential explanation for this limited uptake by the wider community of learning scientists is the high demand of time and human resources that is required for the design, implementation and evaluation of any knowledge community approach.

There is an evident challenge to any application of a knowledge community model in secondary science courses. Knowledge building discourages topic-centered discussions (Bereiter, 2003) in favor of inquiry problems. It requires a long time – potentially up to 6 months – to get the community established sufficiently well that they could identify conceptually inclusive problems that connect multiple topics together Scardamalia (2003). Both of these challenges – the open-ended nature of topics and the time-intensive aspect of community development – are ill suited to secondary science where teachers must make efficient use of curricular time to target a well specified set of science topics. In one of the few published accounts of a KB approach to secondary science (Van Aalst & Chan, 2007) the authors acknowledge making several compromises and falling short of a full implementation of KB.

Considering the challenges that researchers face when studying a knowledge community approach in classrooms, one appealing approach is to develop new models that integrate Web 2.0 technologies with known collaboration scaffolds (Slotta & Peters, 2008). Web 2.0 technologies can provide increased awareness of shared knowledge and help scaffold students in collaboratively constructing new knowledge. This could facilitate the application of a knowledge community approach by researchers in a secondary science context.

## Scaffolded Collaborative Inquiry

Inquiry-based learning (IBL) approaches support students in the process of constructing, refining, and applying their understanding of scientific concepts through question-driven, evidence-based, iterative scientific inquiry processes (Krajcik et al., 2000; Linn et al., 2003).

Collaboration is valued in IBL. However, enacting collaborative inquiry within a classroom is challenging, given the longstanding knowledge transmission culture as well as practical limitations such as time and teacher to student ratio (Singer et al., 2000). The quality of collaborative work may suffer without sufficient support, as a result of students' inexperience with fundamental processes such as pooling ideas, distinguishing between valid and invalid ideas, and settling conflicts (Krajcik et al., 2000; Linn & Slotta, 2009).

Technological and pedagogical scaffolds have been used to relieve complexities of collaborative inquiry. For example, Singer et al. (2000) propose that by having small groups of students share their inquiry questions, plans and findings with the whole class, the students can improve the quality of their work.

Different forms of collaborative design patterns that support inquiry have been advanced by researchers partly driven by the increasing interest in fostering communities of inquiry in classroom. These patterns can provide detailed description of the steps required for students to complete a collaborative task, such as reciprocal teaching (Palincsar & Brown, 1984). While this approach suggests great promise, structuring collaborative inquiry in classrooms remains an important topic of research.

### Knowledge Community and Inquiry (KCI): A Hybrid Model

Knowledge Community and Inquiry (KCI) is a hybrid pedagogical model that integrates the IBL and knowledge community approaches to facilitate a culture of collective inquiry in classrooms. KCI is being developed to make knowledge communities more accessible to secondary science instruction, including a new role for scaffolded inquiry activities that put the collaborative knowledge-base to use in targeting specific learning goals.

Design principles for KCI science curricula are: (1) *Establishing a shared knowledge-base.* Students identify issues of interest related to the topic of the unit and develop knowledge artifacts—e.g. annotated resources or web pages—that embody the emerging themes to direct subsequent inquiry investigation. (2) *Collaborative inquiry.* Once teachers identify issues deemed important by the class that are suitable for inquiry, they work with researchers and technologists to co-design scripted-inquiry activities that draw upon the shared knowledge-base as a resource for the inquiry activities. Decisions about the kind of scripts to be used in the inquiry activity are made based on the topic of discussion and learning goals identified by the teacher. (3) *Continuing to develop the knowledge-base.* The class is engaged in ongoing knowledge integration, connecting their new ideas and understandings to the existing knowledge-base, which can be used in subsequent inquiry activities. (4) *Technology Scaffolds.* Technology plays an essential role in KCI to help students and teachers visualize collective knowledge, to promote every student's contribution to constructing the shared knowledge-base, and to enable students to easily find and retrieve elements from the knowledge base. (5) *Curriculum Co-design.* Designing curriculum that is accessible to teachers is of prime importance to KCI. Any curriculum design must address learning goals held by the teacher, including school district or government curriculum expectations. Moreover, the design team must accommodate the requirements of conventional assessments and must therefore design activities that yield knowledge artifacts that can be graded for content expectations and that prepare students for existing exams.

## Research Questions

The following questions guided this study:

- How will an improved KCI model that relates collaborative inquiry activities to shared knowledge-base impact students' contribution to the classroom knowledge community?
- What are the characteristics of students' contributions to the community knowledge-base?
- How do students' contributions to the knowledge-base relate to their understanding of scientific concepts?
- How do pedagogical and technological scaffolds impact students' contribution to the knowledge-base?

## Methods

Design-based methods bring researchers and classroom teachers together to create innovative approaches to problems of practical and theoretical value, allowing them to recursively examine the interactions among multiple design elements to gradually improve the design (Collins, Joseph, & Bielaczyc, 2004). Edelson (2002) identifies two different objectives for design research: Theory testing and theory development. Considering the purpose of this study, which is to understand how a KCI-inspired technology-integrated curriculum can support scientific inquiry while promoting a knowledge community in science classrooms, a design-based method, geared toward theory development, is appropriate.

### Participants and Settings

Participants of this research are one science teacher and 42 grade ten students in two classes (A and B). The participating teacher has been actively involved in previous research on the KCI model. Following a co-design approach (Penuel, Roschelle, & Shechtman, 2007), the teacher joined the research team in August 2008 for curriculum design, implementation, and evaluation. These meetings continued throughout the unit to refine materials and address unforeseen situations.

### Co-designed curriculum

The curriculum started with an initial knowledge construction activity where students identified useful resources about climate change then annotated and added them under relevant category in a designated wiki page. To promote students' metacognitive awareness of knowledge community, a discussion about "philosophy of science" was also conducted in the first stage of the curriculum. Students then identified important issues related to climate change in Canada that were used to guide small group collaborations. Collaborative inquiry activities included small-group, cross-group, and whole class collaborations. Students formed regional groups to investigate the effects of climate change within each region with regards to three Science dimensions (Greenhouse effect, Weather and wind patterns and Ocean circulation). Within the regional groups, students were assigned to one of the following "specialist roles:" primary industries, secondary industries, environmental activism and tourism industries. For the final project, specialists from all regions joined new advisory groups to synthesize the community's knowledge and provide national guidelines for reducing the adverse affects of global climate change in Canada. All groups were formed across two sessions of the class so that students collaborated with some peers asynchronously. For this iteration, technological infrastructure consisted of a password-protected wiki with sophisticated grouping and permission capabilities, in addition to a collaborative concept mapping software application. Laptops with wireless Internet connections were used for collaborative work. The teacher was highly competent in using technology for teaching.

## Data Sources

Multiple data sources were used to capture the richness and complexity of the innovation:

- Contents of student-created wiki pages provided a measure of the quality of students' ideas and their collaborative knowledge construction.
- Pre-post semi-structured interviews with the teacher and selected students provided deeper insight into the teacher's and the students' perspectives.
- Classroom observation field-notes were collected to help identify patterns of practice and to corroborate data about students' engagement in collaborative activities.
- Formal classroom assessments done by the teacher were used as one factor in examining the quality of students work.

## Analysis and coding scheme

Comprehensive analysis of data is still underway, and indeed a second iteration of the design study is presently in data collection stages. In this paper, we focus on a portion of findings that examines the effect of this co-designed unit on: a) The symmetry of students' contribution within their small groups (ie, in collaborative inquiry); and b) the quality of students' contribution to the wiki. The objective is to understand how curricular activities promote distributed cognitive responsibility and construction of cohesive and usable shared knowledge. For this purpose, we analyze the progress of two small groups: those who created the Ontario and the Prairies regional wiki pages. For any wiki page analyzed, an elaborate process of evaluating every successive version of the page supports our inferences about collaboration patterns and the growth of ideas within the page. Our unit of analysis was an independently meaningful and cohesive piece of text added to the wiki pages. A unit could vary from one sentence to a multi-sentence paragraph.

<u>Wiki development actions:</u> To measure individual group members' contribution to the wiki pages, we reviewed the actions taken by students while they were editing a page, following a constant comparative method, to identify recurring types of edit actions. This analysis sheds light on the extent to which KCI curriculum promoted distributed cognitive responsibility by revealing whether students go beyond their personal contribution to act on what has been added to their group work by other group members. The resulting categories for coding were: Initiating a new section or major idea (initiate); editing content previously added by self (edit self); editing content previously added by another group member (edit other); making link to another wiki pages (link to page); correcting grammatical mistakes or formatting (minor edits).

<u>Quality of wiki contributions:</u> individual group members' contribution to the wiki pages in progress was coded using a four-point scale: Unelaborated facts, elaborated facts, unelaborated explanations, elaborated explanations (adapted from Hakkarainen, 2003). Another category of "regulatory notes" was added subsequently to capture wiki entries concerned with planning, monitoring, or reflecting on the group's overall progress or specific developments.

# Findings and Discussion

Two regional groups, "Ontario" and "Prairies," were selected for this detailed analysis based on the teachers' assessment of their collaborative work. Groups were formed by the teacher and balanced to have both higher and lower achieving students. The mean for the two classes' final unit marking was 71.16 (SD=15.6). The Ontario and Prairies groups were selected as cases because they respectively received lowest and highest marks among seven regional groups. We hypothesized that close examination of these two groups' collaborative inquiry and knowledge construction activity could reveal challenges and advantages in our curriculum design that could guide our design of more coherent collaborative group work and further refine the KCI model.

The Ontario group consisted of six students: Cara, Cheryl, Sam, and Chris (from classroom A) and Sara and Mike (from classroom B). Of these, three scored higher than class mean and three scored lower. The Prairies group had five students: Wendy, Thomas, and Adriana (from classroom A) and Sheila and Mat (from classroom B). Adriana was absent for the final month of the curriculum, which directly affected her contribution to wiki development. We decided to disqualify her contributions from our analysis, which were generally quite minor. In this group, one student scored lower than class mean in the unit and the remaining three scored higher. On average, students in the Prairies group (mean=79.64) received higher final grades than the Ontario group (mean=72.9) in the unit.

## Collaborative Knowledge Construction

Using the "wiki development actions" scheme, we coded all versions of the main wiki page for each region, and all the pages linked to each page that had at least one author from the respective group. Each time a student clicked on the "edit" button, this was counted as one edit session. The total number of such edits was tallied as "Total Page Edit" score. Note that it was possible that during and edit session a student took several actions. Students' individual scores and group means for each item in the coding scheme were then calculated. Figure 1 shows the distribution of edit actions among students from the two groups.

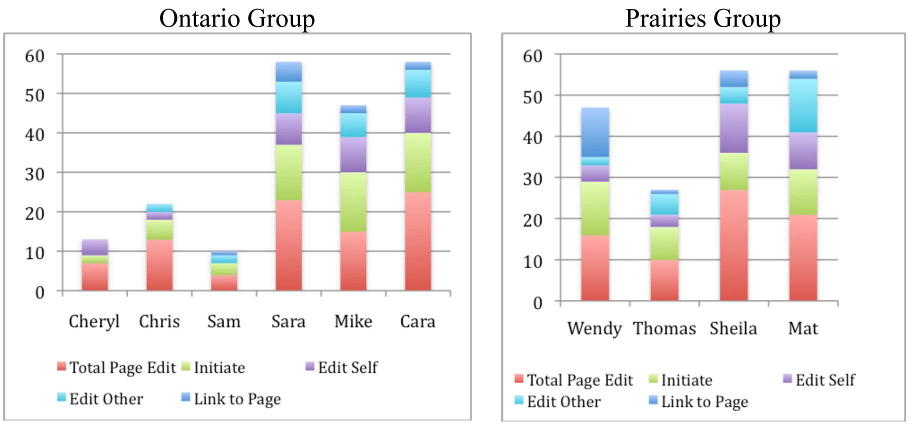



Figure 1. Students' individual scores for "wiki development actions" in Ontario and Prairies groups

The Ontario group was observed by one of the authors throughout the unit implementation. In classroom A, Cara took a leading role, dividing the work among group members. She also took on a high proportion of the work, making 25% of the edits herself (in group size of 6). In classroom B, Mike and Sara made close to the same number of edits, but unlike the classroom A students who often communicated about their collaborations, Mike and Sara only briefly discussed their collaboration in the beginning of the unit, then worked independently afterwards. Sam had the lowest "Total Page Edit" count from the Ontario group. He was present in all group work sessions, and often engaged his group members in relevant discussions. However, Sam typically only searched for information or edited the wiki when he was asked to do so by other group members.

The Prairies group, although not directly observed, demonstrated a rather more evenly distributed participation in wiki development actions. As we will discuss in the following section of Findings, this group took a more calculated approach to planning and monitoring their collaborative work.

To examine the overall performance of the two groups regarding wiki development, we compared the mean score for each action, as illustrated in Figure 2. The Prairies group showed overall higher means in all wiki development actions. Moreover, students in both groups were more likely to initiate new section of text than to edit the existing material added by other group members or linking relevant pages to their wiki page. Figure 2 shows numerical values of means for comparison. These differences were not statistically significant, presumably due to small group sizes.

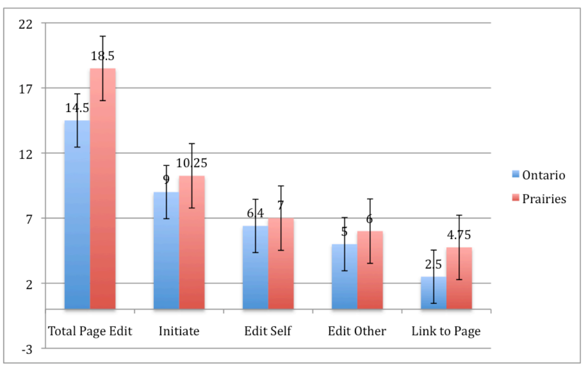



Figure 2. Comparing the means of wiki development actions for Ontario and Prairies groups

Of great interest to this research was the impact of the co-designed curriculum on students' learning. Of the four wiki development actions two provided some measure of students' progressively developing the wiki as a collectively owned artifact rather than adding independent chunks of text to it: Edit Other and Link to Page. We thus calculated a "Collaborative Knowledge Construction" score for each student by adding up her/his "Edit Other" and "Link to Page" scores. A correlation coefficient of 0.83 (Figure 3) suggested a strong positive relation between students' Collaborative Knowledge Construction score and their final grade in the climate change unit.

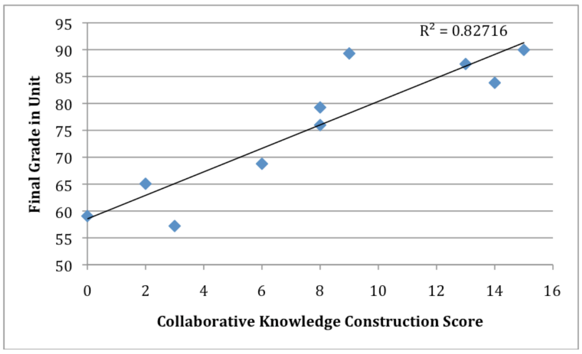



Figure 3. Relation between students' collaborative knowledge construction score in Ontario and Prairies group with their final grade in the unit

## Cohesiveness of co-constructed knowledge objects

To cover the scientific issues that needed to be addressed in the inquiry activity, students took a divide and conquer approach to group work. Since the wiki platform we used did not allow for concurrent edits, students decided to create sub-pages to elaborate on their chosen part of the activity. We were interested to know if the students finally integrated their knowledge into a cohesive final report.

Content analysis of the wiki pages revealed that students in Ontario group occasionally ignored existing wiki contents and added somewhat repetitive content to selected pages. For example Sara and Mike added the following sentences to one wiki page:

"Ontario is Canada's biggest polluter. Over half of Canada's population lives in Ontario, accounting for more waste products and pollutant." (Added by Mike on Dec 18 2008)

"Ontario has the largest population in Canada, and thus Ontario's GHG emissions will be the highest." (Added by Sara on Jan 8 2009)

Mike's entry was a standalone paragraph on top of the page about "Climate Change Issues" and by the final version of the page this paragraph was redundant. Sara's entry was an introduction into a section on "Carbon Emissions" and was well connected to the rest of the section. Nevertheless, the first entry was never deleted or merged with relevant content to make the page more cohesive and less disconnected.

We also noticed contradictory wiki entries in Ontario group. Factors that impacted the amount of released green house gases were discussed in different pages and by different students. The discussion about this issue started by Chris in the main page where initially green house gases were a sub-section of the main page"

Chris : Only about 29% of Ontario's energy is taken from fossil fuel sources- the rest is nuclear and hydroelectric (as Ontario is full of rivers that are used for hydroelectric energy)."

Meanwhile, Sara took the responsibility for developing the "Ontario Greenhouse Gases" page by turning the sub-section heading into a new page which included the following entry related to hydroelectric power:

"Sarah : "Contrary to popular belief, hydroelectric power may sometimes have more GHG emissions than the use of fossil fuels."

The link to this page was directly located above Chris' contribution on December 18 2008. Later, Chris added another entry right under his December 18 2009 entry:

Chris "Since hydroelectricity generates very little air pollution, greenhouse cases in Ontario are very rarely from power generation "

Although his contribution fell under the green houses gases topic, he did not add it to the "Ontario Greenhouse Gases" page. We assume that Chris did not read this page, otherwise he would have noticed the apparent contradiction between Sara's and his own contributions and the two of them would resolve it with more detailed evidence.

In the Prairies group, students showed higher awareness of their shared content. They often communicated with each other about their contributions through meta-comments. For example in one occasion Mat copied Thomas' contribution to its relevant sub-page, "Greenhouse Gases", edited it to decrease the redundancy, and weaved it into his own sentences previously added to that page. However he did not delete the text on the main page that was added by Thomas. Mat also left a comment on the main Prairies wiki page:

Matt: "I put it in the green house gases page and split up your points into jot note form, and got rid of some that I've already added there"

An overall comparison of the two series of wiki pages developed by Prairies and Ontario students revealed a lack of cohesiveness where contents were compartmentalized into subpages without being outlined and summarized in an introduction or a summary section. This could be a flaw of our curriculum design, as we did not require the students to synthesize their wiki work as the collaborative activity concluded. It could be inferred that students closely followed the instruction given to them, which did not directly asked for a cohesive presentation of collected information.

## Quality of contributions

Wiki pages of Ontario group contained more content than those of Prairies group. This could be attributed to the fact that there were more students in the Ontario group. Table 1 shows the quality of students' individual contributions to Prairies and Ontario wiki pages. Students in the Prairies group made extensive use of regulatory notes to communicate division of labor and state of work among themselves. In the Ontario group, on the other hand, regulatory notes were used only in one occasion to ask for all group members to work harder.

Table1. Quality of students' individual contributions to Prairies and Ontario wiki pages

| | | Unelaborated Fact | Elaborated Fact | Unelaborated Explanation | Elaborated Explanation | Regulatory Notes |
|---|---|---|---|---|---|---|
| Prairies Group | Wendy | 7 | 6 | 6 | 2 | 2 |
| | Thomas | 6 | 4 | 4 | 1 | 6 |
| | Sheila | 9 | 1 | | 1 | 2 |
| Ontario Group | Mat | 14 | 6 | 19 | 9 | 1 |
| | Sara | 12 | 10 | 9 | 6 | 3 |
| | Mike | 19 | 9 | 10 | 10 | 2 |
| | Chris | 2 | 4 | 2 | 1 | 0 |
| | Sam | 3 | 6 | 2 | 0 | 1 |
| | Cara | 7 | 8 | 11 | 9 | 0 |
| | Cheryl | 8 | 10 | 4 | 0 | 0 |

An "overall quality score" was calculated for each individual by adding the scores s/he obtained in each of the five categories of content quality. The relation between overall quality scores and final unit grades also resulted in a positive correlation with a coefficient of 0.72 suggesting that students whose wiki entries are more elaborated are likely to achieve higher marks in the unit.

## Conclusion and Significance

In this paper, we shared our interim findings regarding symmetric participation in collective inquiry from the first iteration of a design-based research that examines a hybrid KCI model. This new model encourages even distribution of cognitive responsibility towards building a deep collective understanding of science topics and aspires that students would benefit from each other's contributions to the shared knowledge. Findings shared here, will guide our curriculum design efforts for subsequent iteration of the study in 209-2010.

From the beginning, we expected to observe non-symmetric participation in the inquiry activities both in small group and in whole-class settings. It was our intention, thus, to study the patterns of participation in-situ and design appropriate technological and pedagogical scaffolds to support collaborative activities that maximize the benefits that a knowledge community can offer for all students. Using a wiki as collaboration venue allowed us to study and analyze the process of developing a shared knowledge base through inquiry by disclosing individual student's patterns of participation and their contributions to the final group product.

Careful examination of wiki development actions revealed that students in both groups were more likely to initiate semantically and conceptually new section than to edit the existing material or linking relevant pages to their wiki page (refer to figure for numerical value of means). Upon examining the assessment rubric given to students for their group work, we realized that the students were not explicitly encouraged to act on material added by their group members. A level 4 group contribution was described as "….work in class built on and extended on previous material on wiki" and "material added improved wiki". Yet known characteristics of knowledge communities require students not only to focus on the quality of their individual contributions but also monitor and act on material added by other group members. For these students it was their first experience to make a shared wiki page. "Edit Other" and "Link to Page" wiki development actions in Prairies group showed higher mean which could be due to Prairies students' self-realization that a contribution made to a shared wiki became the intellectual property of all group members. This remains only as speculation as we did not probe these students motivation for group work.

Our findings suggest that students need to be supported in understanding inquiry processes and in actively participating in a collaborative activity since uneven participation may lead to achievement differences. For our next iteration, we are proposing to include two kinds of scaffolds: A reflective self-assessment (e.g. White & Frederickson, 1998), both for small groups and for individuals, to assist students in explicitly monitoring and improving their participation in co-constructing a shared knowledge base; and a participation gauge scaffold to visualize contribution level per student.

Further analysis of the current data will look into the connections that students made between among knowledge artifacts shared by the class.

## Scientific Significance

Developing knowledge communities in classrooms has proved difficult (Lehrer & Schauble, 2006) and inquiry-based learning approaches still seek strategies to promote collective knowledge construction within classrooms (Linn, Clark, & Slotta, 2003). We build on recent research that integrates structured collaborative patterns to complement collective knowledge construction and inquiry process to alleviate these challenges (Hmelo-Silver & Barrows, 2008).

## References

Bielaczyc, K., & Collins, A. (1999) Learning communities in classrooms: A reconceptualization of educational practice. In C.M. Reigeluth (Ed.), Instructional Design Theories and Models, Vol. II. (pp 269-292). Mahwah NJ: Lawrence Erlbaum Associates.

Brown, A. L. (1997). Transforming schools into communities of thinking and learning about serious matters. American Psychologist, 52(4),399-413.

Brown, A. L., & Campione, J. D. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), Classroom lessons: Integrating cognitive theory and classroom practice (pp. 229–270). Cambridge, MA: MIT Press.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. Journal of the Learning Sciences, 13(1):15-42.

Edelson, D. C. (2002). Commentary: Design research: What we learn when we engage in design. Journal of the Learning Sciences, 11(1), 105-121.

Hakkarainen, K. (2003). Progressive inquiry in a computer-supported biology class. Journal of Research in Science Teaching, 40, 1072–1088.

Hmelo-Silver, C., & Barrows, H. (2008). Facilitating collaborative knowledge building. Cognition and Instruction, 26(1), 48-94.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – a conceptual analysis. Educational Psychology Review, 18,159-185.

Krajcik, J., Blumenfeld, B., Marx, R., & Soloway, E. (2000). Instructional, curricular, and technological supports for inquiry in science classrooms. In Minstell, J. and Van Zee, E., editors, Inquiry into inquiry: Science learning and Teaching, (pp. 283-315). American Association for the Advancement of Science Press, Washington, D.C.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. In W. Damon, R. Lerner, K. Anne Renninger, & I. E. Sigel, (Eds,), Handbook of Child Psychology, Sixth Edition, Volume Four: Child Psychology in Practice. (pp 153-196). Hoboken, NJ: John Wiley & Sons.

Linn, M. C., Clark, D., and Slotta, J. D. (2003). Wise design for knowledge integration. Science Education, 87(4), 517-538.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension- fostering and comprehension-monitoring activities. Cognition and Instruction, 1,117-175.

Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. Research and Practice in Technology Enhanced Learning, 2(1), 51-74.

Scardamalia, M. (2003). Knowledge society network (ksn): Toward an expert society for democratizing knowledge. Journal of Distance Education, 17, 63-66.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences (pp. 97-118). New York: Cambridge University Press.

Sherin, M. G., Mendez, E. P., & Louis, D. A. (2004). A discipline apart: the challenges of 'fostering a community of learners' in a mathematics classroom. Journal of Curriculum Studies, 36(2), 207-232.

Singer, J., Marx, R. W., Krajcik, J., & Chambers, J. C. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. Educational Psychologist, 35(3), 165-178.

Slotta, J. D., & Peters, V. (2008). A blended model for knowledge communities: Embedding scaffolded inquiry. Proceedings of the International Conference of the Learning Sciences. Utrecht.

Slotta, J. D., & Linn, M. C. (2009). WISE Science: Inquiry and the Internet in the science classroom. Teachers College Press.

van Aalst, J., & Chan, C. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. Journal of the Learning Sciences, 16(2), 175-220.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. Cognition and Instruction, 16(1), 3-118.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. Journal of the Learning Sciences, 18(1), 7-44.

# The Construction, Refinement, and Early Validation of the Equipartitioning Learning Trajectory

Jere Confrey, Alan Maloney, North Carolina State University, Raleigh, NC 27606
jere_confrey@ncsu.edu, alan_maloney@ncsu.edu

**Abstract.** Equipartitioning, a foundational construct of rational number, is introduced in the form of a learning trajectory. Its development is linked to synthesis of research, the conduct of clinical interviews, and the development of related diagnostic assessments. Refinement and validation of a learning trajectory is described as an iterative empirical process that can serve as a prototype for design of other learning trajectories.

There is perhaps no more important conceptual area in mathematics education than rational number reasoning. The basis of the multiplicative conceptual field (Vergnaud 1983, 1996), rational number reasoning underpins algebra, higher mathematical reasoning, and the quantitative competence required in science. Failure to develop robust rational number construct reasoning and skills in elementary and middle school continues to plague American students. Rational number reasoning is complex, and mastery represents cognitive synthesis—understanding, distinguishing among, modeling, and interweaving a remarkable assortment of distinct yet closely related concepts over many years.

More than four decades of research in rational number topics have provided us with a robust outline of the concepts that comprise rational number reasoning. As described over years of research in the Rational Number Project, and recently outlined by Confrey and colleagues (Confrey 2008; Confrey et al. 2009), rational number reasoning encompasses the constructs of equipartitioning, multiplication and division, ratio, rate and proportion, scaling, length and area measurement, fractions, decimals and percents. That research and summary of rational number strongly suggests that while Rational Number Reasoning is complex, it will yield to a Learning Trajectories analysis (Confrey et al. 2009).

Once learning trajectories are developed, it is assumed that they can be deployed as a framework for curricular innovations and for strengthening teachers' understanding of and instruction in rational number. When coupled with innovative cognitive diagnostic psychometric approaches, learning trajectories may provide a robust means for teachers to assess students' cognitive progress.

"Learning trajectory" is a term with roots in Simon's (1995) definition of "hypothetical learning trajectory," a teacher-conjectured path of student learning from beginning to end of an instructional episode. Learning trajectories are also described in the research literature as predictable sequences of constructs that capture how knowledge progresses from novice to more sophisticated levels of understanding. Clements and Sarama (2004) suggested somewhat more predictable, larger-scale structures, "descriptions of children's thinking as they learn to achieve specific goals in a mathematical domain, and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a developmental progression of levels of thinking" (p. 83). Science educators define *learning progressions* as "empirically-grounded and testable hypotheses about how students' understanding of, and ability to use, core scientific concepts and explanations and related scientific practices grow and become more sophisticated over time, with appropriate instruction" (Corcoran et al. 2009, p. 8). Other views of learning trajectories include the process of developing key conceptual structures (Catley et al. 2004) and that instruction should trace a prospective *developmental* (Brown & Campione 1996) or *conceptual*, *corridor* (Confrey 2006; Lehrer & Schauble 2006), spanning grades and ages, with central concepts introduced early in the school experience and progressively refined, elaborated, and extended. We use a refined definition of "learning trajectory" (or "learning progression"):

> Learning Trajectory: A researcher-conjectured, empirically supported description of the ordered network of constructs a student encounters through instruction (i.e. activities, tasks, tools, forms of interaction and methods of evaluation), in order to move from informal ideas, through successive refinements of representation, articulation, and reflection, towards increasingly complex concepts over time. (Confrey 2008; Confrey, Maloney, Nguyen et al. 2008; Confrey et al. 2009).

While learning trajectories do not dictate a hard-and-fast order in which topics must be learned for students to be successful, they permit specification, at an appropriate and actionable level of detail, of ideas students need to know during the development and evolution of a given concept over time. This information can be embedded systematically in state curricular standards (Confrey & Maloney in press), can underlie curricular interventions with varying degrees of explicitness, and guide teachers in responding to and anticipating student thinking. Learning trajectories may form the backbone of an assessment strategy that can both document

p. 1/1

student progress and identify individual and subgroup weaknesses that can be specifically and efficiently addressed by teachers.

The DELTA project (Diagnostic E-Learning Trajectories Approach to Rational Number Reasoning) at North Carolina State University has the multiple objectives of creating learning trajectories for up to seven distinct strands of rational number constructs, validating them through empirical study, and then designing diagnostic assessments, to assist teachers in targeting their instruction to more effectively support students' cognitive progression through the key rational number constructs across grades K-8.

## Methodology for construction, refinement, and early validation of Equipartitioning learning trajectory.

To construct a learning trajectory, an exhaustive collection, review, and synthesis is conducted of mathematics education and cognitive psychology research literature pertaining to topics in rational number. Synthesis demands careful attention to making sense of related studies, explain higher order relations in complex cumulative findings, or explain old findings by postulating new concepts (Cooper 1998). Once the initial learning trajectory is identified, additional open-ended semi-structured clinical interviews (Piaget 1976; Opper 1977) are conducted. In the case of equipartitioning, these interviews were conducted with 42 urban and suburban students in grades pre-K to 6, and lasted 30 to 60 minutes each. Interviews presented participants with various sequences of specific tasks (varying the numbers of items and people in the fair sharing activities) and followed up with particular questions designed to elicit children's explanation of strategies they used to accomplish the tasks, their justification for their solutions to the tasks, and to probe for multiple strategies for accomplishing each particular version of each equipartitioning task. The interviews permit filling in gaps in the learning trajectory.

The next step in the process[1] is to develop paper-and-pencil assessment items, which requires further clarification of the levels of the trajectory, and piloting of the items. Simultaneously, think-alouds are conducted to determine the extent to which student responses on paper express the range of student thinking elicited by the item. In the DELTA research, 14 items were piloted with 95 2$^{nd}$-grade students, and 14 think-aloud interviews were conducted (Wilson 2009). Based on the synthesis work and the interviews, outcome spaces--the range of observed cognitive behaviors--are summarized for each level of the learning trajectory.

Field-testing follows the development of items corresponding to each level of the learning trajectory. In the equipartitioning research, 125 items were created and administered to approximately 5000 students, grades K-7, across 4 North Carolina school districts. All the students in these grades in 7 elementary and 3 middle schools completed the field tests.

Item-scoring rubrics are then developed in a two-stage process: a) small teams examine all responses to an item, draft a rubric, and identify a range of student response exemplars; b) the team presents the rubric and exemplars to the entire research team, at which time the correspondence of the responses to the learning trajectory levels is evaluated. For the equipartitioning LT, rubrics comprised a 3- to 6-level scale, depending on item format, that accommodated the range of sophistication of item responses, from non-response (or unintelligible) to unequivocally correct or complete. Responses lacking sufficient evidence to be categorized were also coded. Other codes were created to facilitate collection of data on particular strategies and misconceptions.

## Defining Equipartitioning. Using cases to build an initial learning trajectory.

A broad variety of mathematics education and cognitive psychology literature bearing on as many readily identifiable aspects of rational number reasoning as possible was assembled into a comprehensive database (approximately 650 publications). The intent was to assemble an exhaustive collection of peer-reviewed journals, edited books and chapters, and conference proceedings (http://gismo.fi.ncsu.edu/database) (Confrey 2008; Confrey, Maloney, & Nguyen 2008). In examining this diverse database for studies of children's early understanding of multiplication and division, fraction, and ratio, to identify evidence of common or earliest cognitive roots of the concepts in rational number reasoning, it was noted that investigators across a diverse range of differently-oriented studies had asked participants in their studies to solve fair sharing activities. These authors had investigated a variety of mathematical and cognitive topics, including one-to-one correspondence (Frydman & Bryant 1988), partitioning (Pothier & Sawada 1983), early fraction understanding (ref), unitizing (Lamon 1996), the partitive fraction quotient construct (Charles & Nason 2000) (Toluk & Middleton 2003), early ratio understanding (Streefland 1991; Confrey & Scarano 1995; Confrey & Lachance 2000), and aspects of fair sharing itself (Hunting & Davis 1991), and the cognitive relationship between fair sharing and counting competency in young children (Pepper 1991; Pepper & Hunting 1998). Many of the authors had observed that children's reasoning in fair sharing activities followed predictable pathways of increasing sophistication. Early on, Pothier and Sawada (1983) explicitly recognized connections between partitioning and rational number ideas, and identified number theoretic properties and properties of particular simple geometric shapes as important components of children's developing reasoning. Students performing

p. 2/2

these fair sharing activities often gave evidence of reasoning about phenomena predicted by the splitting conjecture (Confrey 1988; Confrey & Scarano 1995).

These fair sharing activities possessed the common feature of generating equal-sized groups or parts, an objective initially named by the research team as "partitioning." The partitioning in these cases, however, is distinct from the kind of partitioning identified in part-part-whole number sense. *Equipartitioning,* as it has been named is partitioning with the specific objective of generating equal-sized groups. We explicitly distinguish the term "equipartitioning" from "breaking," "fracturing" or "segmenting" in which the goal is not necessarily the creation of equal-sized groups, and from the general mathematical term "partitioning." The definition of equipartitioning is as follows:

> Equipartitioning: cognitive behaviors that have the goal of producing equal-sized groups (from collections) or equal-sized parts (from continuous wholes), or equal-sized combinations of wholes and parts, such as is typically encountered by children initially in constructing "fair shares" for each of a set of individuals (Confrey et al. 2009).

Four distinct cases of fair sharing activities were identified as a means to unify equipartitioning (Confrey, Maloney, Nguyen et al. 2008). These four cases comprised the original framework of the Equipartitioning learning trajectory (EqPart LT). The cases are listed here, each with an example of one of the familiar contexts:

Case A: sharing a collection of $m$ objects fairly among $p$ people, where $p$ is a factor of $m$. Example: 15 coins of treasure shared fairly among 3 pirates;

Case B: sharing a single (dissectible) whole fairly among $p$ people. Example: 1 rectangular birthday cake shared fairly among 4 people.

Case C: sharing multiple continuous wholes ($m$) fairly among $p$ people, where $m < p$, and producing $m/p$ as a proper fraction. Example: 3 circular cakes shared fairly among 4 people.

Case D: sharing multiple continuous wholes ($m$) fairly among $p$ people, where $m > p$, and producing $m/p$ as an improper fraction or a mixed number. Example: 9 pizzas shared fairly among 4 people.

The initial learning trajectory was essentially a simple-to-complex ordering of the cases A-B-C-D, and accounted for a number of previous findings. For instance, studies of one-to-one correspondence had established that dealing was a fundamental strategy for fair sharing of collections. Studies of partitioning and splitting had already shown that 2-splits of collections or of single continuous whole objects were readily accomplished by very young (3-4 years old) pre-literate children (Hunting & Sharpley 1991). It had also been established that students mastered different levels of splits in a distinct order, but that the order of mastery did not follow a counting sequence: students consistently accomplish 2- then 4- and commonly 8-splits before accomplishing other split-values, and that 3-splits are very difficult, especially on circular shapes. It was also clear that non-even splits of circles presented far more difficulty for students than splits based on powers of 2 (Pothier & Sawada 1983; Confrey 1997).

### Iterative development of the Equipartitioning learning trajectory.

Articulating and refining the EqPart LT is an iterative process of research synthesis and empirical investigation involving clinical and think-aloud interviews, assessment item response analysis, and rubric development. Previous studies used activities of fair sharing collections of discrete items or single wholes with younger children (pre-K through grade 3), and the more complex tasks of fair sharing multiple continuous items with older elementary and middle school-aged children. However, no systematic examination of the progression of children's reasoning across a wide range of ages and across the entire set of four equipartitioning cases had been conducted. To systematically investigate children's mathematical reasoning in the context of fair sharing across all four cases, semi-structured clinical interviews were conducted with children grades pre-K through 6.

The clinical interviews were designed to elicit a range of student reasoning, including strategies, representations, mathematical practices, emergent properties, and misconceptions. Each interview probed a variety of task parameters, i.e. evaluating students' ability to accomplish tasks of fairly sharing various numbers of items for various numbers of people, and to describe the results of reversing the activity, i.e. reassembling the shares. For each combination of items and people among whom the item(s) were to be shared, students were also asked to verify that each person had received a fair share, justify how s/he knew this, to name the fair share, to identify how much of the collection or item(s) each person received, to share the collection or item(s) in at least one other way, and to identify one $n^{th}$ of the total collection or item(s). Additional questions, designed to elicit student reasoning about the consequences of, for instance, changing the numbers of persons or items at the conclusion of one task, were also employed. A typical interview task, for instance Case A, involves providing a child with a set of plastic "gold" coins, telling her that two pirates have discovered a treasure, and they want to share the treasure fairly. The interviewer then proceeds to ask the child several questions as described above. Once the child had worked through the entire interview for the first number of pirates (2), the interviewer then repeated the questions with a different number of pirates (for example, 3, then 4, or vice versa). Through analysis of these interviews, the EqPart LT was further specified into a more-or-less linear progression of levels, systematically moving through the cases. The interviews were intentionally open-ended: interviewers probed student reasoning when the students' responses or comments presented the opportunity, in order to take note of unexpected, original, or elaborated student behaviors and reasoning.

A critical development was the growing recognition of the importance of properties in explaining how students' cognitive reasoning progressed. While each of the cases required its own variety of strategies, across the cases particular properties concerning sets of tasks emerged, such as compensation (the recognition that increases/decreases in the number of persons sharing results in decreases/increases in share size), composition of splits (acting both vertically and horizontally on rectangles to produce multiplicative numbers of shares), transitivity (the recognition of equivalence of non-congruent fair shares generated from equal-sized wholes), distributivity of equipartitioning over breaking or fracturing of multiple wholes, etc. Developing proficiency in these "emergent properties" led to a re-design of the learning trajectory into a two-dimensional matrix.

Thus, early on in the development of the EqPart LT, the iterative process employed to refine it highlighted the need to differentiate between variations in the task parameters and the increased cognitive sophistication for the levels of the learning trajectory. By this phase of the work, it was clear that the learning trajectory could not be characterized as a simple linear progression of levels of behaviors. The EqPart LT was at this point represented in two dimensions: the "proficiency levels" of the concepts and behaviors on the vertical dimension, and the task parameters on the horizontal dimension. Not all task parameters were applicable to each concept level (for instance, "equipartitioning collections" only pertained to collections of discrete items, not whole dissectible shapes), nor was every concept or behavior relevant to every task parameter (for instance, composing splits on multiple wholes does not apply to single wholes).

In spring 2009, approximately 120 assessment items with individually narrowed focus, covering all relevant combinations of concepts or behavior level and task parameter, were composed and organized into 33 distinct overlapping forms for 4 grade bands: 10 forms of 6 items each for grades K-1 and 2-3, 8 forms of 8 items each for grades 4-5, and 5 forms of 8 items each for grades 6-7. Individual forms were designed for completion within a single class period, to minimally disrupt teachers' normal instructional schedules. The tests were administered to the entire student populations of 7 elementary and 3 middle schools in a rural, an urban, and two mixed suburban/rural North Carolina school districts. The research group examined all responses to each field-tested item in developing the rubrics. An example of an item and its associated rubric is shown in Figures 1 and 2.

Eight family members want to share a whole blueberry pie fairly. This is a drawing of the blueberry pie. Use your pencil to show how the family can share the whole pie fairly.

The blueberry pie

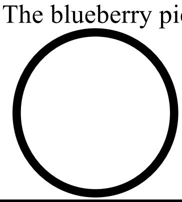
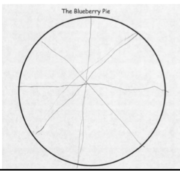
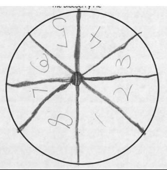
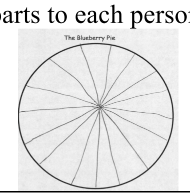
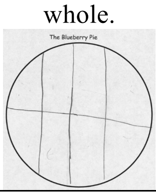
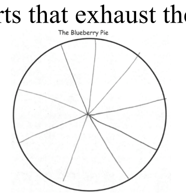


Figure 1. Item (Row 3, Task Class D) of Equipartitioning LT matrix (size reduced from original)

| Level | Descriptions and Exemplars | | |
|---|---|---|---|
| 2 | 2A. Exhausts whole and creates 8 fair shares.   | | 2B. Creates $8k$ equal-sized parts and allocates $k$ parts to each person. Description: Creates 16 equal sized parts and allocates 2 parts to each person. [No exemplar in field test data] |
| 1 | 1C. Creates $8k$ equal-sized parts but does not clearly allocate $k$ parts to each person.  | 1D. Creates 8 unequal-sized parts and exhausts whole.  | 1E. Creates more/fewer than 8 equal-sized parts that exhaust the whole.  |

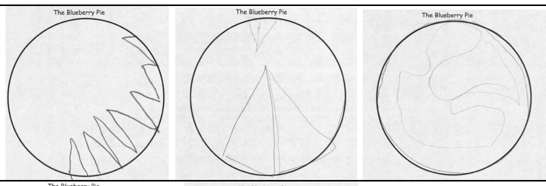
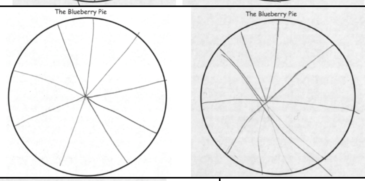
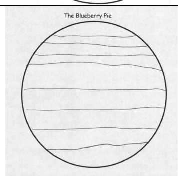
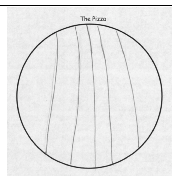
| | | | | |
|---|---|---|---|---|
| 0 | Incorrect or unintelligible. |  | | |
| | 0I. Insufficient evidence. |  | | |
| | Special Code O: Creates *n+1* parts instead of *n* parts (# of cuts versus # of parts) (Here, makes 8 cuts (9 parts) when sharing among 8 people |  | Special Code P: Creates parallel cuts to try to fairly split a circle (shape misconception). |  |

Figure 2. Rubric, based on student responses, for the item in Figure 1.

This phase moves the work closer to the aim of deploying diagnostic assessments aligned with the progression of concepts in the learning trajectory. Each item's responses are examined in detail to determine whether the item elicits a consistent and interpretable range of student responses that justifies its continued usefulness for inferring student cognitive understanding. Items needing revision to improve clarity or specificity are identified at this point, as are those that should be redesigned or even discarded.

Table 1. Equipartitioning Learning Trajectory Matrix. Proficiency levels form the vertical dimension, in the left column of the table. Task classes, listed along the top row, form the horizontal dimension.

| **Equipartitioning Learning Trajectory Matrix (Grades K-8)** | | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** | **L** | **M** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Task Classes→ Proficiency Levels | | Collections | 2-split (Rect/Circle) | $2^n$ split (Rect) | $2^n$ split (Circle) | Even split (Rect) | Odd split (Rect) | Even split (Circle) | Odd split (Circle) | Arbitrary integer split | $p = n + 1$; $p = n-1$ | $p$ is odd, and $n = 2^i$ | $p >> n$, $p$ close to $n$ | all $p$, all $n$ (integers) |
| 16 | *Generalize*: $a$ among $b = a/b$ | | | | | | | | | | | | | |
| 15 | *Distributive property,* multiple wholes | | | | | | | | | | | | | |
| 14 | *Direct-, Inverse- and Co-variation* | | | | | | | | | | | | | |
| 13 | *Compositions* of splits, multiple wholes | | | | | | | | | | | | | |
| 12 | Equipartition *multiple wholes* | | | | | | | | | | | | | |
| 11 | Assert *Continuity principle* | | | | | | | | | | | | | |
| 10 | *Transitivity* arguments | | | | | | | | | | | | | |
| 9 | Redistribution of shares (*quantitative*) | | | | | | | | | | | | | |
| 8 | Factor-based changes (*quantitative*) | | | | | | | | | | | | | |
| 7 | *Compositions* of splits; *factor-pairs* | | | | | | | | | | | | | |
| 6 | *Qualitative compensation* | | | | | | | | | | | | | |
| 5 | *Re-assemble*: $n$ times as much | | | | | | | | | | | | | |
| 4 | *Name* a share with respect to referent unit | | | | | | | | | | | | | |
| 3 | *Justify* the results of equipartitioning | | | | | | | | | | | | | |
| 2 | *Equipartition single wholes* | | | | | | | | | | | | | |
| 1 | *Equipartition Collections* | | | | | | | | | | | | | |

p. 5/5

Item analysis and rubric development also serve a bidirectional role in the empirical process of refining and validating the learning trajectory. The outcome spaces for individual items help the research group determine whether individual items are consistent with and likely to be useful in inferring student progress with respect to the conceptual level from which the items were originally designed, and whether the item should be assigned to a different conceptual level. Conversely, and perhaps more important, the analysis of outcome spaces provides a form of test of the learning trajectory itself. The outcome spaces for different items of different cognitive levels are closely examined to ascertain whether the relative complexity and sophistication of reasoning evidenced by the students across items of different levels are consistent with the hypothesis represented by the levels of the EqPart LT, that of progression of student cognition from less sophisticated to more sophisticated.

## Levels and Task Parameters of the Equipartitioning Learning Trajectory

The Equipartitioning Learning Trajectory is represented in a two-dimensional matrix (Table 1) comprised of the progression of proficiency levels along the vertical axis (listed from bottom to top), and the task classes along the horizontal axis (listed from left to right, roughly in order of case A to case C/D tasks). The proficiency levels and task classes are organized to accommodate the following behavioral and cognitive components of student reasoning for each equipartitioning case.

- The extent to which students successfully accomplish equipartitioning.
- Strategies and representations students use to accomplish the tasks, including multiple approaches
- Evidence of mathematical reasoning practices, including naming and justifying
- Emergent mathematical properties
- Predictable patterns of errors: misconceptions and critical barriers

## Discussion

The analysis to date of clinical interviews and assessment item responses permit us to further support several major conclusions identified previously by Confrey and colleagues (Confrey 2008; Confrey et al. 2009):

- The EqPart LT is robust for children across the entire age range K-7. Older children in this age range, by and large, exhibit the same progression of reasoning about the tasks, strategies, and emergent properties as do younger children.
- Pothier and Sawada's (1983) conclusion, that number theoretic properties and geometric qualities are important in (equi)partitioning activities, is strongly supported.
- These foundations of rational number concepts do not require addition or subtraction operations, and can be developed in school in parallel to the development of those operations, rather than subsequent to them.
- Through Equipartitioning, (partitive) division cognitively precedes multiplication.
- Multiplication as re-assembly of equipartitioned collections or wholes can readily be conceptually developed instead of and earlier than it is typically developed as repeated addition, and that this would greatly strengthen children's subsequent multiplicative reasoning. We believe that deriving multiplication from repeated addition, as currently done in most curriculum, limits the development of children's multiplicative reasoning, constraining it inappropriately to additive reasoning, when it should be developed as a distinct but parallel conceptual reasoning field;
- Equipartitioning of multiple whole items leads in the higher levels of the EqPart LT to the distribution of equipartitioning over breaking or fracturing of sets or items, offering a foundation for the distributive property that can precede the numeric and algorithmic treatment of the property, and thus strengthen student flexibility with this critical concept that is essential for algebra, and which bridges additive and multiplicative reasoning.
- Behaviors, mathematical reasoning practices, and emergent mathematical properties that are elicited through equipartitioning of collections and continuous wholes, by children as young as 5 years old, anticipate the following rational number concepts:
  - Ratio, through two-dimensional 'many-to-one' numerical relationships, unit ratios, and ratio units;
  - Fraction-as-number, through the re-unitizing of fair shares to a many-as-one definition of the resulting fair share, unit fraction (students readily identify a single share of $n$ fair shares as "$1/n^{th}$ of the whole), and possibly the partitive fraction quotient construct;
  - $a/b$-as-Operator, through the development of "$1/n^{th}$ of" and "$n$ (times) as many" in naming fair shares and identifying the referent units for the fair shares resulting from equipartitioning.

## Conclusions

The contributions of the current work include the definition and elaboration of a learning trajectory for Equipartitioning, a rational number construct that integrates previously disparate cases. This was done through a rigorous and systematic process of research synthesis, clinical interviews, and extensive analysis of student item responses from a large body of assessment items developed to align with the case-based analysis of fair sharing activities.

The development of assessment items and analysis of student item responses with respect to the learning trajectory levels is an empirical process that serves to validate the ordering and specific content of levels of the learning trajectory, on the one hand, and the opportunity to correct and re-align those levels with student responses.

Our work on learning trajectories in rational number reasoning represents a large-scale design problem. The work to date on the Equipartitioning learning trajectory provides a prototype for operationalizing the design of rational number learning trajectories as frameworks for instruction and assessment. Additional study of equipartitioning is underway, directed at development of corresponding diagnostic assessments. Further validation steps include the field testing of additional items, further elaboration of the EqPart LT proficiency levels, and the application of different statistical measurement models to prototype assessments.

A significant opportunity for further work is the construction of supplemental curriculum materials based on the Equipartitioning learning trajectory. We recall that we define learning trajectories as specifically involving an "ordered network of constructs a student encounters *through instruction*…" We note that Equipartitioning is not part of any curriculum of which we are aware. It is notable that older students who have not become proficient in equipartitioning exhibit misconceptions that could be remedied by systematic introduction to equipartitioning. This suggests that if equipartitioning were embedded in elementary mathematics curriculum, understanding of rational number could be accelerated in major ways.

## Endnotes

(1) The process outlined here is informed by the "BEAR method" (Wilson 2005), adapted to link to our synthesis and interview protocols.

## References

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education*. Mahwah, NJ: Lawrence Earlbaum Associates.

Catley, K., Lehrer, R., & Reiser, B. (2004). *Tracing a prospective learning progression for developing understanding of evolution*. Paper presented at the Committee on Test Design for K-12 Science Achievement. from http://www7.nationalacademies.org/bota/evolution.pdf

Charles, K., & Nason, R. (2000). Young Children's Paritioning Strategies. *Educational Studies in Mathematics, 43*(2), 191-221.

Clements, D. H., & Sarama, J. (2004). Learning Trajectories in Mathematics Education. *Mathematical Thinking and Learning, 6*(2), 81-89.

Confrey, J. (1988). Multiplication and Splitting: Their Role in Understanding Exponential Functions. In M. Behr, C. LaCompagne & M. Wheeler (Eds.), *Proceedings of the Tenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 250-259). DeKalb, IL: Northern Illinois University Press.

Confrey, J. (1997). *Student Voice in Examining "Splitting" as an Approach to Ratio, Proportions and Fractions* Paper presented at the 19th Annual Meeting of the International Group for the Psychology of Mathematics Education, Recife, Brazil.

Confrey, J. (2006). The evolution of design studies as methodology. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 135-152). New York: Cambridge University Press.

Confrey, J. (2008). *A Synthesis of the Research On Rational Number Reasoning: A Learning Progressions Approach to Synthesis.* Paper presented at the 11th International Congress of Mathematics Instruction, Monterrey, Mexico.

Confrey, J., & Lachance, A. (2000). Transformative teaching experiments through conjecture-driven research design. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of Research Design in Mathematics and Science Education* (pp. 231-265). Mahwah, NJ: Lawrence Erlbaum Associates.

Confrey, J., Maloney, A., Nguyen, K., Mojica, G., & Myers, M. (2009). *Equipartitioning/Splitting as a Foundation of Rational Number Reasoning Using Learning Trajecories.* Paper presented at the 33rd Conference of the International Group for the Psychology of Mathematics Education, Thessaloniki, Greece.

Confrey, J., Maloney, A., Nguyen, K., Wilson, P. H., & Mojica, G. (2008). *Synthesizing Research on Rational Number Reasoning*. Paper presented at the Research PreSession of the Annual Meeting of the National Council of Teachers of Mathematics.

Confrey, J., Maloney, A., & Nguyen, K. H. (2008). Rational Number Reasoning Database, from http://gismo.fi.ncsu.edu/index.php?module=Search

Confrey, J., & Maloney, A. P. (in press). Linking Mathematics Standards to to Learning Trajectories: A Case Study from North Carolina, 2008-09. In F. Mosher, J. Confrey & P. Daro (Eds.). New York, NY: Columbia University Press.
Confrey, J., & Scarano, G. H. (1995, October 21-24). *Splitting reexamined: Results from a three-year longitudinal study of children in grades three to five.* Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Columbus, OH.
Cooper, H. (1998). *Synthesizing Research*. Thousand Oaks, California: SAGE Publications, Inc.
Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning Progressions in Science: An Evidence-based Approach to Reform*. New York: Center on Continuous Instructional Improvement Teachers College–Columbia University.
Frydman, O., & Bryant, P. (1988). Sharing and the Understanding of Number Equivalence by Young Children. *Cognitive Development, 3*, 323-339.
Hunting, R. P., & Davis, G. (Eds.). (1991). *Early Fraction Learning*. New York: Springer-Verlag.
Hunting, R. P., & Sharpley, C. F. (1991). Pre-fraction concepts of preschoolers. In R. P. Hunting & G. Davis (Eds.), *Early fraction learning* (pp. 9-26). New York: Springer-Verlag.
Lamon, S. J. (1996). The Development of Unitizing: Its Role in Children's Partitioning Strategies. *Journal for Research in Mathematics Education, 27*(2), 170-193.
Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 371-388). Cambridge: University Press.
Opper, S. (1977). Piaget's clinical method. *Journal of children's Mathematical Behavior, 5*, 90-107.
Pepper, K. L. (1991). Preschoolers Knowledge of Counting and Sharing in Discrete Quantity Settings. In R. Hunting & G. Davis (Eds.), *Early Fraction Learning* (pp. 103-129). Heidelberg: Springer-Verlag.
Pepper, K. L., & Hunting, R. P. (1998). Preschoolers' Counting and Sharing. *Journal for Research in Mathematics Education, 29*(2), 164-183.
Piaget, J. (1976). *The Child's Conception of the World*. Totowa, NJ: Littlefield, Adams, & Co.
Pothier, Y., & Sawada, D. (1983). Partitioning: The Emergence of Rational Number Ideas in Young Children. *Journal for Research in Mathematics Education, 14*(4), 307-3317.
Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114-145.
Streefland, L. (1991). *Fractions in realistic mathematics education: A paradigm of developmental research*. Dordrecht, the Netherlands: Kluwer Academic Publishers.
Toluk, Z., & Middleton, J. A. (2003). The Development of Children's Understanding of the Quotient: A Teaching Experiment. *unpublished*.
Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematical concepts and processes* (pp. 127-174). New York: Academic.
Vergnaud, G. (1996). The theory of conceptual fields. In L. Steffe & P. Nesher (Eds.), *Theories of mathematical learning* (pp. 219–239). Mahwah, NJ: Lawrence Erlbaum.
Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
Wilson, P. H. (2009). *Teachers' Uses of a Learning Trajectory for Equipartitioning.* North Carolina State University, Raleigh, NC.

## Acknowledgments

The work presented here was supported by NSF Awards DRL-0758151 and DRL-0733272. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors wish to acknowledge ongoing contributions to the DELTA project by P. H. Wilson, K. H. Nguyen, G. Mojica, M. Myers, R. Pescosolido, C. Edgington, A. Franklin, and P. Larios.

# Finding Transactive Contributions in Whole Group Classroom Discussions

Hua Ai, Marietta Sionti, Yi-Chia Wang, Carolyn Penstein Rosé, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, {iamhuaai,marietsionti}@gmail.com, {yichiaw,cprose}@cs.cmu.edu

**Abstract:** Automatic analysis of discussion logs from studies of computer supported collaborative learning has become more prevalent as technology for machine learning and text mining have become more powerful and usable. We extend this effort to analysis of transcripts of whole group classroom discussions. We argue in favor of adapting a transactivity based framework developed for analysis of collaborative learning discussions for use in the context of whole class discussions. Results from an evaluation of this technology applied to transcripts from 3 class periods of middle school math demonstrate promising results, specifically Cohen's Kappa of .69 in comparison with human annotators for identifying transactive contributions and .68 for identifying which prior utterance a transactive contribution relates to. Potential applications for teacher professional development, automatic assessment of learning processes, and real time support for classroom discussion facilitators is discussed. Implications for theory building related to learning in classroom discussions is also addressed.

**Keywords:** Corpus analysis, classroom discourse, automatic text processing techniques, argumentation

## Introduction

In recent years the learning sciences community has seen a growing body of work towards automatic analysis of conversational interactions in learning contexts (Soller & Lesgold, 2000; Cakir et al., 2005; Erkins & Janssen, 2006; Rosé et al., 2008). Much of this work has taken place within the computer supported collaborative learning community (Wang et al., 2007; Kumar et al., 2007; Kumar et al., 2009), as part of an effort to develop dynamic support for collaborative learning that is capable of triggering interventions in a way that is responsive to what is happening within the conversation. In this paper we present an extension of automatic conversation analysis technology into a new area, specifically analysis of whole class group discussions. Our automatic analysis involves a construct referred to as Transactivity (Teasley, 1997; Weinberger & Fischer, 2006; Berkowitz & Gibbs, 1983), which is well studied in the domain of educational psychology and computer supported collaborative learning. Transactive contributions are arguments constructed in such as way as to reference, sometimes described as "operating on", the previously expressed reasoning of self or others.

Transactivity is an important construct to target for automatic analysis technology because of the extent of its prevalence within the literature on collaborative learning process analysis and the body of work that supports its value as a property of discussions for learning (Azimita &Montgomery, 1993; Joshi & Rosé, 2007). The body of work already demonstrating the feasibility of using machine learning technologies to identify transactive conversational moves in small group discussions (Rosé et al., 2008; Joshi & Rosé, 2007) offers hope that it can be extended into this new context. Ideas related to effective patterns of discussion in classroom contexts have evolved within their own separate history from that of the community of researchers studying analysis of collaborative learning interactions. Nevertheless, a growing subcommunity of the classroom discourse community has focused on facilitation strategies for group discussions that have very similar motivations relating to encouraging children to articulate their reasoning and to listen to and respond to the reasoning of others (Chapin & O'Connor, 2004; Resnick et al., 1992; Resnick, Michaels, & O'Connor, in press; Michaels, O'Connor, & Resnick, 2007). Similarly, within the problem based learning community, where discussion groups are smaller, but similarly lead by skilled facilitators, again similar ideas have emerged (Hmelo-Silver & Barrows, 2006). Thus, it is reasonable to expect that similar technology would be capable of identifying transactive moves in whole class discussions. Nevertheless, large group discussions are more complex than those that take place within small groups of two or three participants. Thus, extension of earlier automatic analysis technology to whole group discussion data is a research contribution in its own right.

In the remainder of the paper we describe our adaptation of the Berkowitz and Gibbs transactivity coding scheme for classroom discussion. We then describe experiments towards automatic analysis of transactivity from transcripts of classroom discussions. We then discuss potential applications of this technology for teacher professional development, automatic assessment of learning processes, and real time support for classroom discussion facilitators. Implications for theory building related to learning in classroom discussions is also addressed. We conclude with a discussion of remaining challenges for producing automatic transactivity analyses from classroom discussions.

## Theoretical Framework

This paper presents our preliminary study on automatically analyzing transactivity in a classroom setting. This study is part of a larger effort related to building conversational agents that can assess the level of transactivity in ongoing discussions on the fly and respond accordingly. This work is based on the assumption that an assessment of level of transactivity in an in-progress discussion would be valuable information for facilitators to use in deciding what is needed from them to keep the conversation on a productive path in their role as orchestrators of group talk. This assumption is based on investigations of valuable conversational contributions in learning contexts that have been conducted both within communities exploring the cognitive foundations of group learning as well as the sociocultural community. Regardless of the theoretical framework, the same ideas have surfaced pointing to the value of transactivity or other very similar constructs as properties of conversation for learning. It should be noted that we are not arguing that this transactivity assessment is sufficient for facilitators to decide how to proceed, only that it provides useful insight into the level of productive engagement each student in the group is or has been exhibiting over the course of the conversation. Our work focuses on discussions involving young children, specifically middle school children learning math, although transactivity as a construct is not specific to math, and thus its definition does not refer to any specific math knowledge. Nevertheless, in order for the construct to be applied in its most valid sense, as we will see below, knowledge of math would be required in order to fully identify all of the connections between reasoning displays that students are making. Because of this, the approach we describe in the next section of this paper will be seen as necessarily a simplification of this construct. Nevertheless, the results show promise that although the annotations are being applied through automatic means, the reliability in comparison with human coding is high. Methodological implications of such simplifications are discussed at length by Rosé and colleagues (2008) in connection with automatic analysis of small group collaborative learning discussions.

The purpose of this section is to argue for the value of this transactivity analysis in connection with prior work in the learning sciences. For example, one cognitive justification for the value of transactive conversational behavior is its connection with cognitive conflict (Piaget, 1985). While we do not argue that transactivity is synonymous with cognitive conflict, we do argue that where transactive conversational moves highlight differences between the mental models of collaborating students, these moves provide opportunities for students to reflect on differences in their own mental models, which may then lead to cognitive conflict. One can argue that a major cognitive benefit of collaborative learning is that when students bring differing perspectives to a problem-solving situation, the interaction causes the participants to consider questions that might not have occurred to them otherwise. This stimulus could cause them to identify gaps in their understanding, which they would then be in a position to address. This type of cognitive conflict has the potential to lead to productive shifts in student understanding. It has the potential to elicit elaborate explanations from students that are associated with learning (Webb, Nemer, & Zuniga 2002). While transactivity has been valued previously for its role in consensus building discussions (Weinberger & Fischer, 2006), it should be noted that conversations exhibiting transactivity need not end in a common understanding or belief about an issue, and its value for learning can be seen as independent of whether this occurs or not. Furthermore, the value of transactivity is not limited to its potential cognitive benefits. From the sociocultural perspective, based on Vygotsky's seminal work (Vygotsky 1978), we can similarly argue that when students who have different strengths and weaknesses work together, they can provide support for each other that allows them to solve problems that would be just beyond their reach if they were working alone.

The most popular formalization of the construct of transactivity (Berkowitz and Gibbs, 1983) has 18 types of transactive categories, which characterize each child's conversational turn, as long as it is considered an explicit reasoning display. Before considering which of these codes, if any, is appropriate for a contribution, one must first determine whether that contribution constitutes an explicit articulation of reasoning, or at least a reasoning attempt. Beyond this, transacts have been divided in three types: elicitational, representational and operational, which ultimately were reduced to two, incorporating the elicitational in representational (R), which is considered a lower level transact, since it elicits or re-presents another's reasoning. On the contrary operational transacts (O) present a person's new argumentation, which is formed by building on another's contribution. A transact may also combine both types (R/O), because the boarders might be vague in some cases. The other two dimensions of transactive moves are focus and mode. Depending on the primary focus, a transact might be self-oriented (ego, operates on the speaker's own reasoning) or other-oriented (alter, operates on the reasoning of a partner, dyad shared opinion). Mode indicates if the transact was expressed competitively (i.e., the two expressions of reasoning are not consistent with one another) or non-competitively (i.e., the two displays of reasoning are consistent with one another). We have simplified the formulation considerably in this work and simply identify utterances according to whether they do not count as an articulation of reasoning, count as reasoning but not transactive, and transactive reasoning. In the case of transactive utterances, we identify which previous utterance is being related to.

In our formulation, articulation of reasoning by students is the goal, and thus we define what "counts" as a reasoning move. Our formulation of what counts as a reasoning display comes from the Weinberger &

Fischer (2006) notion of what counts as an "epistemic unit", where what they look for is a connection between some detail from a scenario (which in their case is the object of the case study analyses their students are producing in their studies) with a theoretical concept (which in their case comes from the attribution theory framework, which the students are applying to the case studies). When they have seen enough text that they can see in it mention of a case study detail, a theoretical concept, and a connection between the two, they place a segment boundary. Occasionally, a detail from a case study is described, but not in connection with a theoretical concept. Or, a theoretical concept may be mentioned, but not tied to a case study detail. In these cases, the units of text are considered degenerate, not quite counting as an epistemic unit.

We have adapted the notion of an epistemic unit from Weinberger & Fischer, rather than using it the same way Weinberger et al. did in their work, both because the topic of our conversations is very different in nature and because we're working with a much younger group of students. We consider that the basic requirements for a unit of talk to count as a reasoning display is that it has to contain evidence of a connection between some detail from the problem the students are trying to solve and some mathematical idea, which could be a theorem or an idea from an earlier problem they solved that they explicitly mentioned (because it shows evidence of making an abstraction), or some idea from a book that they explicitly mentioned. We would like to distinguish this from just parroting what they have heard. While teachers may use repetition to keep an idea salient within a discussion, which may be a very valuable facilitation move, we would rather not count places where students simply repeat what they have heard as evidence of reasoning. In our current formulation, we are considering the problem that the teacher has set up as what is given. We would like to make a distinction between what is given and what the students contribute beyond that.

So far we have located the instructor somewhat outside of the discussion the students are having, seeing the instructor as stimulus and support for the discussion and not actually part of the discussion. Note that we consider this scaffolding by the instructor as extremely valuable, and even necessary. However, we see it as distinct from the effect that teachers are attempting to achieve in the conversation between the students. Thus, in our current formulation, we are making a distinction between moves that facilitate the drawing out and refining of the kids' reasoning from the actual reasoning attempts. Thus, this paper only focuses on the analysis of the student utterances. As an additional caveat, while it was true also for analyses of adult discussions, it is even more true of these child discussions that we need to allow for displays of incorrect, incomplete, and incoherent reasoning to count as reasoning, as long as in our judgment we can believe an attempt at reasoning was going on. That will necessarily be quite subjective – especially in the case of incoherent explanations. However, we have achieved a human agreement of .68 Kappa using the definitions of articulation of reasoning and what it means to be a transactive utterance.

## Machine Learning Technology

Machine-learning algorithms can learn mappings between a set of input features and a set of output categories. They do this by using statistical techniques to find characteristics of hand-coded "training examples" that exemplify each of the output categories. The goal of the algorithm is to learn rules by generalizing from these examples in such a way that the rules can be applied effectively to new examples. The goal is to replace this low level coding effort by humans rather than to perform an analysis over human coded data as has been done frequently in other work (e.g., Nathan, Kim, and Grant, 2009). In order for the machine learning process to work well, the set of input features provided must be sufficiently expressive, and the training examples must be representative. Typically, machine-learning researchers design a set of input features that they suspect will be expressive enough. At the most superficial level, these input features are simply the words in a document. But many other features are routinely used in a wide range of text-processing applications, such as word collocations and simple patterns involving part of speech tags and lexical features; we will draw from this prior work.

Once candidate input features have been identified, analysts typically hand code a large number of training examples. For example, tools such as the publically available TagHelper tool set (Rosé et al., 2008) have the capability of allowing users to define how texts will be represented and processed by making selections on the GUI interface. In addition to basic text-processing tools such as part-of-speech taggers and stemmers that are used to construct a representation of the text that machine-learning algorithms can work with, a variety of algorithms from toolkits such as Weka (Witten & Frank, 2005) are included in order to provide the means to learn mappings between the input features and the output categories.

## Evaluation

We evaluated the use of machine learning technology for the analysis of transactive conversational moves in three classroom sessions where a teacher worked with a group of kids to elicit their reasoning about integers. Altogether the corpus covers 185 minutes of discussion involving the teacher and the students. The sessions have been transcribed. Altogether, the transcripts contain 707 student turns and 646 teacher turns. In this corpus, the students and the teacher contribute about equally.

## Predicting Student Moves

Student moves were coded as Not Reasoning (NR), Externalization (E), or Transactive (T). Not reasoning is a large category that includes all student utterances that are off task, are related to management (e.g., responds to announcements), are plain answers to a teacher's question, or are simply repeating what had already been articulated. A detailed discussion can be found in (Sionti et al., 2010). Reasoning utterances are grouped into two categories. Externalization refers to students' reasoning that does not connect with any previously displayed reasoning. Transactive utterances are student utterances that display reasoning that connects with previous reasoning displays in some way. For example, in this dialog exerpt below, the first student utterance is tagged as Externalization while the second student utterance is tagged as Transactive.

> Student: I sort of agree but I also disagree. I wanted to say maybe the weight line shows you more because it shows like the materials and it also showed how much they weighed by how far apart they are...
> Teacher: Can you say a bit more about that?
> Student: Like the mineral oil is way heavier than the whatsit [looks at chart], the organic materials, but the mineral oil and the water, the fresh water are way closer. Like you could fit the length of 2 mineral oils in the length of 1 gravel.

Table 1: Student Moves in three classroom sessions

| | T | E | NR |
|---|---|---|---|
| session1 | 75 | 41 | 112 |
| session2 | 77 | 25 | 126 |
| session3 | 90 | 25 | 136 |

In the case of transactive contributions, the prior utterance that the transactive utterance refers to, or operates on, is identified. Table 1 displays counts of student moves (**stuMov**) in each of the classroom sessions. We observe that students produced turns identified as articulations of reasoning (T or E) 47.1% of the time. Thus, a trivial classifier that simply assigns the category of NR, which was the most frequent category, would achieve a percent accuracy of 52.9%, although this corresponds to a Cohen's kappa of 0, which we will treat as our baseline for comparison. In the case of identifying which utterance a transactive utterance is responding to, again a simple heuristic achieves a moderate accuracy. Specifically, when predicting responding student utterance ID (**respondUtt**), we can achieve and accuracy of 54.3% by always predicting the current student turn refers to one utterance back. So again, we will treat this as our baseline, and our experiments will evaluate how much improvement we can achieve with a more sophisticated approach in comparison to these simple baselines.

Table 2: Summary of features used in machine learning experiments.

| Category | Features | Definition |
|---|---|---|
| Lexical | Utt | the string of words appearing in the current student utterance |
| Length | Len | the length of the current student utterance |
| Topic | chageOfTopic | a binary feature that indicates whether the main topic of discussion has shifted in last utterance |
| speaker Info | speaker | a binary feature that indicates whether the current speaker is one student or a group of students |
| | lastSpeaker | A binary features that indicates whether the last speaker is a student, a teacher, or a group of students |
| | changeOfSpeake r | a binary feature that indicates whether the student is also the speaker in the previous student turn |
| LSA scores | LSA1 | the LSA score representing the comparison between the current student turn and the last previous turn |
| | LSA2 | the LSA score representing the comparison between the current student turn and the turn that is two turns back |
| | LSA3 | the LSA score representing the comparison between the current student turn and the turn that is three turns back |

In our work, we built two separate classifiers to predict stuMov and respondUtt. In both cases, we first evaluated different classification algorithms by exploring combinations of different supervised learning algorithms and different sets of features on the session 1 data. Then, to test the generality of the approach we identified as best on this data, we evaluated the best performing approach on the whole corpus.

We used three standard supervised classification algorithms that have shown promising performance in previous work (Rosé et al., 2008), namely Naïve Bayes (NB), support vector machines (SMO), and decision trees (DT). For each classifier, we explored a variety of combinations of features, many of which are made available very simply through TagHelper tools. Table 2 defines the features we used and organizes them into 5 categories. Note that what is identified as the LSA score is calculated using Latent Semantic Analysis (Foltz et al., 1998), in which we calculate the semantic similarity between two utterances using LSA's latent factor approach based on patterns of co-occurrences between words. ChangeOfTopic is the only manual feature that is tagged by a human annotator. Prior work has shown that automatic topic segmentation of conversational data is feasible (Arguello & Rosé, 2006).

In our first set of experiments, we evaluated the alternative machine learning algorithms with different sets of features. The first set of features contained only the lexical features. The next set of features included that in addition to the length feature. The next set included all of those features in addition to the topic feature. The next set included all of those features in addition to the speaker information features. The next set included those plus the LSA features. The final set included all except for the lexical features.

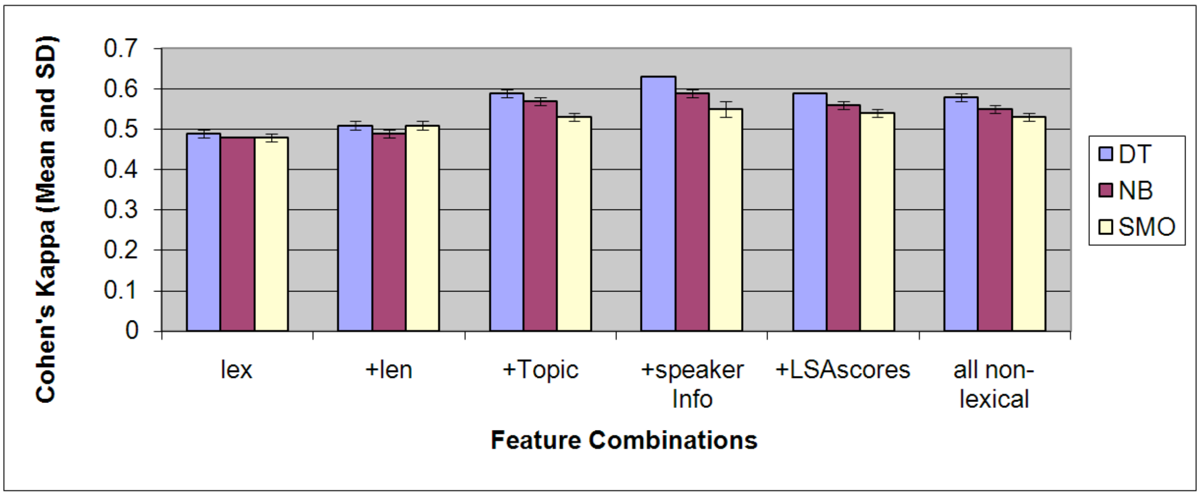



Figure 1. Classification results on predicting stuMov.

Figure 1 presents our results on predicting stuMov in the session 1 data. The bars show the Cohen's kappa and the error bars show the standard error. "+ Feature Category" stands for adding the new category of features in addition to the features that have already been used in the feature sets on its left. In the last set of feature "all non-lexical", we use all non-lexical features. We observe that the decision tree classifier trained on lex+len+changeOfTopic+speakerinfo features give us the best kappa of 0.63. Decision tree learning is also the best classifier across all feature combinations. Thus, it is reasonable to expect that this is a stable finding.

We were not satisfied with the result we were able to achieve with these combinations of features and algorithms. Thus, we conducted an error analysis in order to identify ways in which we could improve our performance. We began by looking for common confusions the best performing approach made between categories. Table 3 displays what is referred to as a confusion matrix, with allows us to see where these common confusions are occurring.

Table 3: confusion matrix of best performing classifier on predicting stuMov

| | predicted NR | predicted E | predicted T |
|---|---|---|---|
| target NR | 99 | 3 | 10 |
| target E | 7 | 19 | 15 |
| target T | 3 | 13 | 59 |

.

We can see that most of the misclassified cases are between "E" and "T". The difference between "E" and "T" is whether a student is expressing his own reasoning or relating to some previously articulated reasoning, which would have typically been contributed by a different speaker. We hypothesized that we can reduce this type of error by taking into account more information about previous contributions where reasoning was articulated. In other words, we want to utilize the predicted student moves on previous turns. As an approximation of this type of approach, we first assign first pass categories to utterances using our three category coding scheme, and then we use these assigned categories as features in order to apply the rules found in Figure 2, which describes a heuristic post-processes stage.

| Rule | Definition |
|---|---|
| 1 | After the teacher changing to a new topic, there shouldn't be T before E (if there is T, change it to an E) |
| 2 | After changing to a new topic, if the same student gives an NR, then teacher asks "why", the next utterance from this student is likely be E, not T |
| 3 | If a student's utterance is an E, teacher ask students to explain, then the next student utterance should be a T |

Figure 2. Heuristic rules for post-processing stuMov predictions.

We implemented theses heuristic rules so that they only apply in cases where the machine learning classifiers make a prediction with low confidence. We define low confidence as confidence scores which are lower than the average confidence scores over all student utterances. Note that although the heuristic rules are manually authored, they can be executed automatically by a computer program to assign stuMov. Also, these rules only reinforce the definition of the three stuMov tags that are not captured by the classifier. Therefore, these rules are not aimed to tailor the automatically predicted stuMovs to a specific training corpus.

With the post-processing, the kappa between the predicted stuMov and the human gold standard on the first session is improved to 0.71 (SD=0.01). We further apply the best classifier and the heuristic rules on the whole corpus of all three sessions and get a kappa of 0.69 (SD=0.01), which shows our approach can be generalized to a larger corpus. This result is encouraging, but we consider it preliminary until more extensive testing on a wider variety of transcripts is conducted to ensure generalizability.

## Predicting the Utterance Responded To

We employed a similar methodology in automating the analysis of which utterance a transactive utterance is responding to. We refer to this categorization as respondUtt. The target prediction is a numerical value. However, the classifiers we are using can only handle target predictions which belong to a finite set of values. Therefore, we define the class value for our prediction task to be 0, 1, 2, 4, and other. 0 stands for the current utterance, in other words, it refers to 0 utterances back, which means the current utterance is an Externalization. In contrast, values such as "1", "2", and "4" stand for Transactive moves that refer to the associated number of contributions back. If the utterance that the current utterance refers to is not 1, 2, or 4 utterances back, it is labeled as "other". We will assign a number to "other" cases in a post-processing stage using heuristic rules.

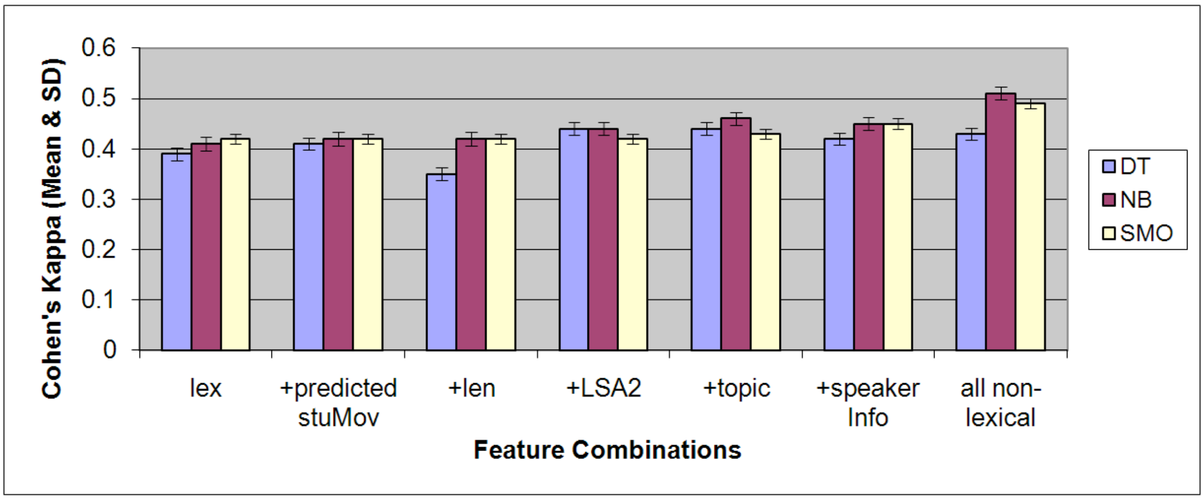



Figure 3. Classification Results on predicting respondUtt

Figure 3 shows the results on predicting respondUtt by the three classifiers in combination with the same sets of features used for predicting the stuMov. We observe that NB using all non-lexical features performs the best in this task, with a kappa of 0.51. Similarly, we add a post-processing step based on heuristic rules shown in Figure 4 using a similar methodology to what we reported in the previous section. By implementing these heuristic rules, the kappa score is improved to 0.69 (SD=0.01). Again, we apply the best performing model with the post-processing step to the whole corpus of all three sessions and get a kappa score of 0.68 (SD=0.01), which offers some verification of the generalizability of our approach.

| Rule | Definition |
|---|---|
| 1 | A student turn with transacMov=T shouldn't point to a student turn with transacMov=NR |
| 2 | A student turn with transacMov=E, respondUtt=0 |
| 3 | A student turn with transacMov=NR, respondUtt=current utterance Num-1 |

Figure 4. Heuristic rules for post-processing respondUtt predictions.

## Discussion

Much work has already been invested in fruitful applications of automatic analysis technology in the area of computer supported collaborative learning (Wang et al., 2007; Kumar et al., 2007). The work presented in this paper points towards a new line of research applying this technology in a classroom context. One potential application of this technology could be for use in teacher professional development, by supporting instructors in reflecting on how their classroom interactions with students have proceeded and how the students are progressing in terms of striving towards articulations of transactive expressions of their reasoning. In support of the students themselves, such technology could be used to track development of a student's argumentation and articulation skills over time. Finally, eventually it may be possible for such technology to provide real time feedback to instructors or students during group discussions in order to stimulate higher levels of transactivity within the discussions. In addition to these practical applications, one can imagine that such technology also holds the potential to speed up the science of investigating the role of patterns of conversational behavior in stimulating valuable social and cognitive processes within classroom contexts. Such arguments have previously been elaborated in connection to research in computer supported collaborative learning (Rosé et al., 2008).

## Conclusions and Current Directions

In this paper we have presented work to date on adapting a transactivity based analysis framework for analysis of classroom discussions and have explored the use of automatic analysis technology in this context. We demonstrated how we used an iterative development methodology to apply and extend available technology to a new analysis task, namely that of analysis of transcripts of whole group classroom discussions. Specifically, we have employed text mining techniques to analysis of transcripts from 3 classroom discussions, and have presented promising results, and positive implications both for practical applications and theory building.

Much work remains to extend this technology before real time analysis of classroom discussions will be practical. Most notably, the work presented in this paper depends upon the availability of transcripts of the classroom interaction. Thus, a considerable effort to record and automatically transcribe classroom discussions in real time using speech recognition technology still requires a substantial and targeted research effort. While advances in the field of speech recognition will undoubtedly be required in order to bring this within practical reach, prior work related to automatic assessment of group processes from speech offers hope that such an effort could be feasible (Gweon et al., 2009).

This work was supported by the Pittsburgh Science of Learning Center, NSF SBE-0836012.

## References

Arguello, J. & Rosé, C. P. (2006). Topic Segmentation of Dialogue, *Proceedings of the NAACL Workshop on Analyzing Conversations in Text and Speech.*

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development*, 2, 202-221.

Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, 29, 399-410.

Cakir, M., Xhafa, F., Zhou, N., & Stahl, G. (2005). Thread-based analysis of patterns of collaborative interaction in chat. *Proceedings of the 12th international conference on Artificial Intelligence in Education (AI-Ed 2005)*, Amsterdam, The Netherlands, pp. 120–127.

Chapin, S., O.Connor, C., & Anderson, N. (2003). Classroom Discussions: Using Math Talk to Help Students Learn, *Math Solutions Publications*: Sausalito, CA.

Erkens, G., & Janssen, J. (2006). Automatic coding of communication in collaboration protocols. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.) *Proceedings of the 7th International Conference of the Learning Sciences (ICLS)* (vol. 2, (pp. 1063–1064)). Mahwah, NJ: Lawrence Erlbaum Associates.

Foltz, P., Kintsch, W., & Landauer, T. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, 25, 285–308.

Gweon, G., Rosé,C. P., Albright, E., Cui, Y. (2007). Evaluating the Effect of Feedback from a CSCL Problem Solving Environment on Learning, Interaction, and Perceived Interdependence, *Proceedings of Computer Supported Collaborative Learning*

Hmelo-Silver, C. E. & Barrows, H. S. (2006). Goals and Strategies of a Problem-based Learning Facilitator. *The Interdisciplinary Journal of Problem Based Learning*, 1(1), pp 21-39.

Joshi, M. & Rose, C. P. (2007). Using transactivity in conversation summarization in educational dialog. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education.*

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial Dialogue as Adaptive Collaborative Learning Support, *Proceedings of Artificial Intelligence in Education*

Kumar, R., Chaudhuri, S., Howley, I., Rosé, C. P. (2009). VMT-Basilica: An Environment for Rapid Prototyping of Collaborative Learning Environments with Dynamic Support, *Community Event Proceedings of Computer Supported Collaborative Learning*

Nathan, M., Kim, S., & Grant, T. (2009). *Instituting Change in Classroom Discourse Structure: Human and Computer-Based Motif Analysis*, Working Paper number 2009-1, Wisconsin Center for Education Research, School of Education, University of Wisconsin, Madison.

Michaels, S., O'Connor, C., & Resnick, L.B. (2007). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education.*

O'Connor, M.C., Michaels, S., & Chapin, S. (2007). Small-scale experimental studies of classroom talk: Seeking local effects in discourse-intensive instruction. *Paper presented at annual meeting of the American Educational Research Association.*

Piaget, J. The *equilibrium of cognitive structures: the central problem of intellectual development*, Chicago University Press, 1985.

Resnick, L. B., Bill, V., & Lesgold, S. (1992). Developing thinking abilities in arithmetic class. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 210-230). London: Routledge.

Resnick, L.B., Michaels, S., & O'Connor, C. (in press). How (well structured) talk builds the mind. In R. Sternberg & D. Preiss (Eds.), *From Genes to Context: New Discoveries about Learning from Educational Research and Their Applications.* New York: Springer.

Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F., (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning, *International Journal of Computer Supported Collaborative Learning* 3(3), pp237-271.

Schwartz, D. (1998). The productive agency that drives collaborative learning. In Dillenbourg, P. (Ed.) *Collaborative learning: Cognitive and computational approaches.* NY: Elsevier Science/Permagon.

Sionti, M., Ai, H., Rosé, C. P., and Resnick, L., A Framework for Analyzing Development of Argumentation through Classroom Discussions, in Niels Pinkwart & Bruce McClaren (Eds.) *Educational Technologies for Teaching Argumentation Skills*, Bentham Science.

Soller, A., & Lesgold, A. (2000). Modeling the Process of Collaborative Learning. Proceedings of the *International Workshop on New Technologies in Collaborative Learning*. Japan: Awaiji–Yumebutai.

Stahl, G. (2006c). *Group cognition: Computer support for building collaborative knowledge.* Cambridge, MA: MIT Press. Retrieved from http://www.cis.drexel.edu/faculty/gerry/mit/.

Suthers, D. (2006). Technology affordances for inter-subjective meaning making: A research agenda for CSCL. *International Journal of Computer Supported Collaborative Learning*, 1, 315-337.

Teasley, S. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, R. Säljö, C. Pontecorvo & B. Burge (Eds.), *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361-384). Berlin: Springer.

Vygotsky, L.S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.

Wang, H. C., Rosé, C.P., Cui, Y., Chang, C. Y, Huang, C. C., Li, T. Y. (2007). Thinking Hard Together: The Long and Short of Collaborative Idea Generation for Scientific Inquiry, *Proceedings of Computer Supported Collaborative Learning*

Webb, N., Nemer, K., & Zuniga, S. (2002). Short Circuits or Superconductors? Effects of Group Composition on High Achieving Students. Science Assessment Performance, *American Educational Research Journal* 39(4), pp943-989.

Weinberger, A. & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*.

Witten, I. H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, second edition, Elsevier: San Francisco.

# Magnetism as a Size Dependent Property: A Cognitive Sequence for Learning about Magnetism as an Introduction to Nanoscale Science for Middle and High School Students

David Sederberg, Lynn A. Bryan, Purdue University, 525 Northwestern Avenue, West Lafayette, IN, 47906.
Email: dsederbe@purdue.edu, labryan@purdue.edu

**Abstract:** In this study, we use evidence from students in grades 10-12 (N = 148) to document the ways in which children learn about and model their conceptual understanding of magnetism. Our model-based design capitalizes on phenomena that are engaging to students, to describe the behavior of magnetic materials across scale, from the familiar to the nanoscale. Data are drawn from written responses to pre- and post-questionnaires, embedded assessments, activity journal pages, and informal interviews, to describe a progression of students' mental models of magnetism. Our goal is to integrate multiple related concepts to describe possible pathways for learning. An implication of this work is that it may be used to inform level-appropriate strategies for both instruction and assessment for the study of magnetism across scale.

## Introduction

Magnetism is long standing staple of science curricula from grades K-12 and beyond. It is a phenomenon that fascinates and interests students of all ages. Yet we know surprisingly little about people's conceptions of magnetism (Hickey & Schibeci, 1999; Maloney, 1985), nor have conceptions of magnetic phenomena been investigated as intensively as other physical phenomena such as electricity, force and heat (Borges & Gilbert, 1998; Erikson, 1994; Hickey & Schibeci, 1999).

Researchers have previously documented students' conceptions related to magnetism such as (a) models of magnetism (Borges & Gilbert, 1998; Constantinou, Raftopoulos, & Spanoudis, 2001; Erikson, 1994); (b) the confusion between magnetism and charge (Borges & Gilbert, 1998; Hickey & Schibeci, 1999; Maloney, 1985); (c) action at a distance (Bar, Zinn, & Rubin, 1997); and (d) the concept of field (Bradamante & Viennot, 2007; Guisasola, Almudi, & Zubimendi, 2004; Guth & Pegg, 1994). Less study has been devoted to the exploration of how students' conceptions about magnetism progress or to document the transitional nature of students' conceptions and mental abstractions as they learn. Nor have students' conceptions of magnetism across scale been investigated. Learning about nanomagnetism, for example, provides a means of not only introducing concepts of nanoscale science to children, but it is a vehicle to reinforce and apply science concepts already existing in the curriculum.

The concepts of magnetism pose a challenge for learners; they require higher levels of cognition and mental imagery than more concrete and tangible concepts (Barrow, 1987; Borges & Gilbert, 1998; Guisasola, et al., 2004). The idea of an object causing a force to be exerted on another without touching, or reasons why an object could be attracted toward either end of a magnet are counterintuitive for children (Constantinou, et al., 2001). Children commonly believe that magnets are electrically charged, and that electrostatic and magnetic interactions are the same (Borges & Gilbert, 1998; Haupt, 1952; Maloney, 1985). Many students may also believe that the magnetic field has a finite boundary (Bar, et al., 1997), or that the field lines are a concrete entity (Guisasola, et al., 2004; Guth & Pegg, 1994). Understanding these concepts requires the ability to construct mental models of abstract concepts, such as spatial orientation, noncontact forces and the particle nature of matter.

The goal of our research is to describe the increasingly sophisticated ways in which students construct mental models about magnetism. We will show how a limited number of key benchmark concepts, through guided inquiry, mental imagery, and metacognitive reflection, can enable students to develop a deeper and more intuitive understanding of magnetism and magnetic phenomena, both in the realm of the familiar, as well as at the nanoscale.

## Theoretical Framework

This work is based on a developmental view of learning "developing relations among a (relatively few) set of core concepts throughout schooling" (Catley, Lehrer & Reiser, 2005, p, 8). These concepts provide targets for learning, accessible to younger students yet still a challenge for in-depth understanding at later grades (Lehrer & Schauble, 2004). Wiggins & McTighe (2006) characterize these core concepts as linchpin ideas - those concepts within a content domain that have enduring value throughout and at multiple levels with the domain, providing conceptual anchors for construction of knowledge and a basis for assessment.

We also draw from the literature on mental models and model-based learning. As a representation of something in the absence of the real thing (Greca & Moriera, 1997) mental models provide a lens through which levels of understanding and cognitive sophistication, relative to a target concept, may be approximated. Model-based learning presumes that through the process of generating, critiquing and revising our mental models we can approach a more coherent and normative understanding (Clement & Steinberg, 2002; Coll, France, & Taylor, 2005; Lehrer & Schauble, 2006; White & Frederiksen, 1998). Models are the dominant form of explanation in science and learning science is to make, revise and justify self-constructed models, not simply to use models posed by others and taken for granted (Lehrer, 2009).

Mental models provide the learner a means to organize concepts in a way to help understand the world or to explain it to others (Harrison & Treagust, 1996). In order to construct a coherent conceptual understanding of scientific principles, learners must be able to formalize, assess and reflect upon, and justify their understanding of scientific concepts as they develop them (Cavicchi, 1997; Clement, 1989; Clement & Steinberg, 2002). In the process of constructing a mental model the learner reduces a phenomenon to the elements most meaningful, selecting "only some parts of the entity and relations between them" to create a personally meaningful representation (Gilbert & Boulter, 1995).

The use of mental models generated and revised by the student while learning and revealed through multiple inscriptions (Latour, 1990) will provide a lens through which both the researcher and learner will be able to access the evolutionary nature and coherence of that cognitive process. The invention and revision of models, as an expression of a form of knowing is characteristic to the natural sciences and "are the defining features of scientific thinking" (Lehrer & Schauble, 2006). Yet, it has been shown that changes in mental models and conceptions do not occur quickly and often require repeated challenges, metacognitive reflection, and multiple mixed models along the way (Nussbaum & Novick, 1982).

One of the impediments to understanding magnetism, especially magnetism at the nanoscale, is the concept of the atom and the particle nature of matter. At the nanoscale the effect that thermal energy has on particle motion is manifested in the behavior of the material. The behavior of these materials is not only relevant to the realm nanoscale science, but serves as a tool for connecting to and learning about the nature of magnetism in general.

Children are aware of the terms atom and molecule and references to both of them as "particles" and yet still believe that matter is continuous or that substances contain molecules rather than substances are composed of molecules (Harrison & Treagust, 2002). A scientific understanding magnetism requires an understanding of the particle nature of matter and the implications of thermal energy and applied forces on the appropriate hierarchy of single or aggregate groups of particles (e.g. atom, domain, magnet).

The overarching goal of this research is to provide an in-depth look at the progression and coherence of students' mental models in learning about magnetism. An important part of our work is the belief that nanoscience and nanotechnology offer an exciting impetus for learning, empowering student learning in the context of new and exciting concepts of emerging s discovery. Furthermore, the introduction of nanoscience into the existing curriculum helps to provide a big picture view of science and conceptual coherence of scientific concepts for students. The specific goals for this study will be to answer the following questions.

1) What is the nature and range of sophistication of grade 10-12 students' initial models of magnetism and magnetic interactions?
2) What common patterns emerge as students critique and revise their mental models of magnetism during the learning process?
3) What themes among students' developing mental models provide coherent explanatory power across scale?

## Method

This study was guided by an orientation in interpretive research (Creswell, 2009), to document the generation and iterative revision of students' mental models of magnetism. We used a quasi-experimental design to compare the construction and progression of mental models of selected concepts of magnetism. Emphasis was placed on students' construction and revision of models, coordinated across multiple magnetic phenomena and scale. Iterative cycles of investigation, reflection and revision were used to provide students the opportunity to revise their models for coherence and explanatory power. The two teachers participating in this study each taught their own magnetism unit, following the lesson format provided by the researchers and, with minor modifications, used the same assessment instruments.

## Participants

The grade 10-12 high school students (N = 148) in this study comprised a non-random sample, solicited from two high schools with which researchers had previously worked as constituents in a nanoscience teacher professional development program. The classes were all high school physics classes; the majority (approximately 95%) of the students in the classes were in the 11th grade (see Table 1).

Table 1: Student Groups Comprising the Sample

| Sample | Description of students | N | Intervention | Code |
|---|---|---|---|---|
| 1 | High school physics (17-18 years) | 66 | Yes | MC |
| 2 | High school physics (17-18 years) | 72 | Yes | PL |

## Intervention

The series of lessons for the intervention will follow a cycle of model-based inquiry in which students will make predictions, conduct investigations, interpret results and apply their revised mental models to new situations. During the cycle of lessons, students will be provided multiple opportunities to reflect on and revise their mental models, and to defend or revise them in light of new and possibly conflicting evidence. The series of lessons which will be employed were designed by the researcher and are based on those concepts believed to be most salient to students constructing a coherent understanding of magnetism and the behavior of magnetic materials across scale.

## Data Sources

The data collected consisted of student responses to pre- and posttest tests, inscriptions from embedded assessments, activity journal pages, collaborative group artifacts and informal interviews. Our goal was not the assessment of effectiveness of an intervention. Rather, we aimed to document the status and growth of students' mental models and explanations of magnetic phenomena through the normal course of classroom instruction by the classroom teacher. Our lessons and assessments for this study consisted of items in the areas of: the concept of magnet; magnetic interactions, magnetic versus electrostatic interactions, magnetic fields, domains, magnetizing and demagnetizing, and magnetism at the nanoscale. The assessment items were paper and pencil open ended response and graphic format.

## Data Analysis

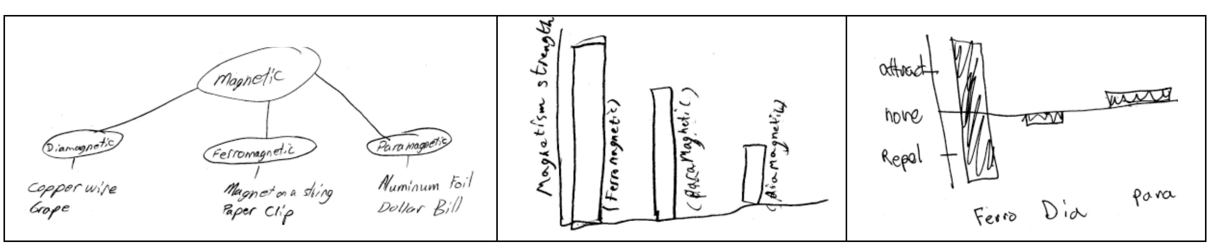
Written responses were transcribed into a spreadsheet for comparison and analysis. Drawings and graphics were coded to fit students' conceptions into categories representing levels of sophistication based on the depth of understanding exhibited, relative to target concepts

# Findings and Discussion

We describe our findings for the following categories: (a) magnetic materials and interactions; (b) magnetic fields; (c) models of magnetization and magnetic domains; and (d) magnetism at the nanoscale.

## Magnetic Materials and Interactions

Students initially conducted an inquiry designed to elicit their models of magnetic interactions and then made a drawing of their own design to interpret their findings. Interactions were typically classified as attractive, but not necessarily repulsive, in various combinations of weak or strong (see Figure 1). Some students focused on the nature of interactions, others by the materials involved.




PL-17 PL-38 PL-6

Figure 1. Classification of Magnetic Interactions

While some students' drawings depict more than one level of interpretation, many lacked the degree of sophistication that would indicate a broad understanding of concepts. Few students included repulsion, for example in their model, despite having used two magnets and despite observing the diamagnetic behavior of a grape. Very few students included the possibility that there could be no magnetic interaction even though some of the materials used were not magnetic (plastic ruler, paper, wood), and students frequently did not make a distinction between strong versus weak interactions. Some students' models indicated common non-normative beliefs, for example that all metals will be attracted to a magnet Table 2 lists categories of responses that might be used to differentiate levels of students' conceptions of magnetic interactions based on their observations with these materials.

Table 2: Concepts of Magnetic Interactions.

| Level | Key features of drawings |
|---|---|
| More sophisticated | Includes magnetic and non-magnetic, attraction and repulsion, strong and weak, includes magnet-magnet interactions; defines types of magnetism |
| | Includes magnetic and non-magnetic, strong and weak, types of magnetism |
| Less sophisticated | Little or no predictive value; reflects only materials used |

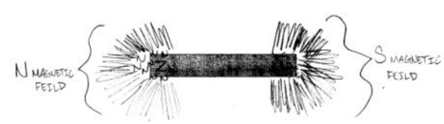
The purpose of this initial lesson was to help students develop an awareness of degrees of magnetic attraction and/or repulsion, how these interactions are observed, and what common materials exhibit these properties. Students re-visited categories of magnetic materials, the degree of response and the ability of some materials to remain magnetized or not, in subsequent lessons.

## Magnetic Fields

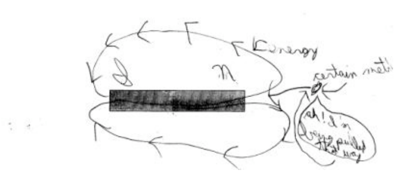
The detection of magnetic fields was an integral part of our lessons, both for investigation and as a tool for detection. The drawing of an iron filing diagram identified a change in the space surrounding a magnet and provided an overall concept of the both the geometry of the field and areas of more and less density. Students also discovered during this activity the three dimensional nature of the field. Mapping the field with a compass provided students access to the directional orientation of the field and identified the compass as a field "detector" for subsequent inquiries, the magnetization and cutting of a wire, and magnet consisting of iron filings in a straw.

The pre- and posttest assessments asked students to draw or write all of the "characteristics" that they think might help to explain how magnets work (see Figure 2.)

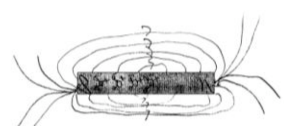
PL-22 Pretest



PL-22 Posttest



PL-50 Pretest



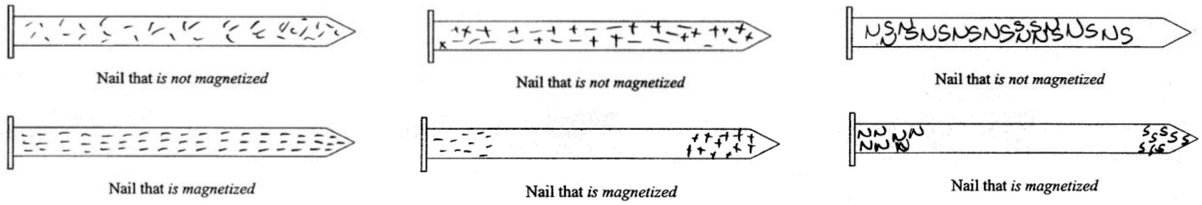
PL-50 Posttest

Figure 2. Pre- and Posttest Drawing of "Characteristics" of a Magnet.

In previous iterations we framed this question more to elicit conceptions specific to magnetic fields. However, we found that leaving the item more open for interpretation produced a broader range of responses. Additionally, we found that the degree to which students might elaborate on their conception of the field helped provide a means for categorizing understanding. The most sophisticated models included enough field lines to convey a greater density at the poles, the symmetric nature of the field, identification of poles and an indication of direction. The role of the field in the alignment of the domains of a ferromagnetic object is a key concept that recurred in subsequent lessons.

## Magnetization and Magnetic Domains

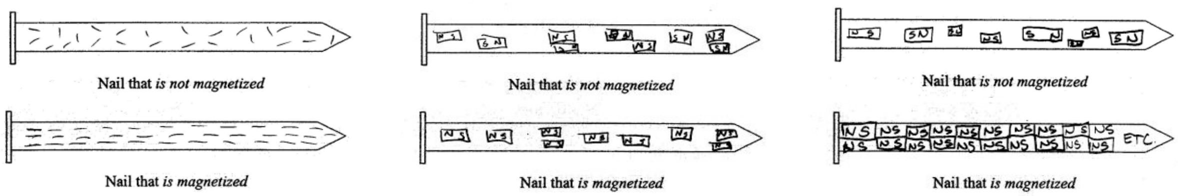
We approached the concept of domains with a stepwise progression of investigations to help students construct, and provide us access to, increasingly sophisticated mental models. At the end of the first investigation, categories of magnetism, we asked students to speculate how a magnetized nail might differ from a non-magnetized nail if they were able to "see inside." These models provide a means of eliciting students' mental models of how magnetized and non-magnetized materials might differ (see Figure 3).




PL-62 PL-14 PL-22

Figure 3. Pre-instruction Models of Magnetized and Non-magnetized Nails

These models provided a beginning for the development of the concept of domains, fundamental in providing the conceptual framework to understand how a nail, for example, can be attracted to either end of a magnet, why an object can be magnetized, or why nanoscale materials do not stay magnetized. At key times throughout the lessons, students referred to their initial models, confronting and revising them based on new evidence and their changing conceptions. Posttest responses for the same students are shown in Figure 4.

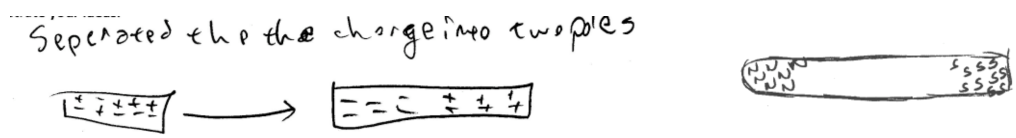



PL-62 PL-14 PL-22

Figure 4. Post-instruction Models of Magnetized and Non-magnetized Nails

These inscriptions demonstrate that while PL-22 and PL-14 had begun to incorporate the concept of domains into their conceptual framework, PL-62's mental model had not changed. PL-22's model would also indicate that the entire nail was comprised of domains, while PL-14's model gives the impression of domains embedded in another material.

One of the several ways in which we made the concept of domain accessible to students was through a "straw magnet." Student made their own magnet by filling a soda straw with iron filings and then magnetizing and de-magnetizing their magnet, observing the change in field with a compass. While students generally began to associate being magnetized with a state of conditional alignment, we found that many still held on to their concept of charge (see figure 5).




PL-14 PL-22

Figure 5. Magnetizing a Straw Magnet

The two inscriptions in Figure 5 are good examples of models between the pre- and post-instructional models in Figures 3 and 4. Both of these models show an increased sophistication from the initial model, yet are refined further during the remainder of the lessons, as the inscriptions in Figure 4 demonstrate.

After the straw magnet lesson, students from one group create a drawing for what they believed would happen when a nail is approached by either end of a magnet (see Figure 6). The other group responded to a similar question with a written response (see Table 3).

Figure 6. How a Nail is Attracted to Either Side of a Magnet (MC-21).

The drawings by student MC-21 are a reasonable model of the alignment and reverse alignment of the domains in the nail, subject to the field of the magnet. This student aligns the domains in the nail relative to the magnet and indicates the action of the field.

Table 3: How a Paper Clip is Attracted to Either Side of a Magnet

| Level | Key concepts |
|---|---|
| More sophisticated | The magnet magnetizes the paper clip, which means it causes the domains within it to be lines up. The paper clip can be attracted to both ends of the magnet because both ends will line up the domains. Therefore, one end of the magnet will line up the domains in one direction and the other end will line them up in the other direction. (MC-66). |
| Less sophisticated | The domains of the paper clip will shift according to what side of the magnet is presented allowing it to pick up the paper clip (MC-67). |

The more sophisticated of the responses alluded to the alignment of domains, yet none of the students in our sample referred to the concept the field in the action; the level of sophistication appeared to stall with the magnet causing the effect. This is however neither surprising nor discouraging, given the limited time frame for instruction and the abstract nature of the interaction.

## Magnetism at the Nanoscale

The magnetic properties of nanoscale materials is grounded in two key concepts: (1) the dominance of the behavior of a particle by the larger number of atoms on the surface of the particle relative to atoms in the interior compared to microscale and macroscale objects, and (2) thermal effects which exceed the tendency for nanoscale domains to remain aligned in the absence of a magnetic field. The investigation and explanation of the behavior exhibited by a magnetic fluid (ferrofluid), ties together all of the concepts presented in the preceding lessons. Figure 7 shows two student inscriptions, representative of two levels of sophistication, applying observations from the nanoscale material investigation with the concepts they learned.

PL-1 PL-47

Figure 7. Response of a Magnetic Fluid to a Magnet.

Inscription PL-1 correctly shows the gathering of the magnetic fluid near the pole of the magnet underneath, but lacks the conceptual detail of the field lines and the corresponding "lining up" of the magnetic particles. The second inscription shows correctly oriented field lines with the magnetic particles conforming to a pattern similar to what students had observed with iron filings.

Students were also asked to explain why nanomagnetic particles do not remain magnetized when larger samples of the same materials do. The rationale behind the question is to elicit students' ability to translate their mental models of magnetization, field, domains, and the effect of thermal energy and particle size, to the behavior of the magnetic fluid (See table 4). We found that, even at the end of the series of lessons, there remained a continued perception of electrostatic charge being a factor in the magnetic behavior of the nanoscale particles among some students' conceptions. The two following responses illustrate, "There is too much surface to keep the charges magnetized and therefore tend to lose charge quickly" (MC-52); "The charges have a bunch of space to move around and don't want to stay in one spot to stay magnetized" (MC-6).

Table 4: Why Nanosize Particles Do Not Stay Magnetized

| Level | Examples of student responses |
|---|---|
| Most sophisticated | These nanoscale particles are constantly moving around, causing the domains to become unaligned at any moment. If the domains are unaligned, the particle cannot stay magnetized (MC-4). |
| Less sophisticated | They don't have enough atoms on the inside to make up domains to stay lined up to keep magnetized (MC-15). |

While the responses in Table 4 hint at aspects of thermal energy and the impact of surface, relative to volume, students may remain "captured" by their intuitive knowledge (Cheng & Brown, 2007), reverting back to that intuition rather than using a still undeveloped but more sophisticated model.

## Conclusion

This research represents a pilot study. Our goal was to begin to document a range and progression of sophistication of students' mental models of magnetism, revealed through multiple and frequent inscriptions. We believe that through model-based learning, mentoring students to construct, metacognitively reflect on, and reform their mental models, learners will approach a more coherent and normative understanding of concepts of science. This study serves as a starting point for further research that aims to develop cognitively grounded and research-based cohesive physical science curricula organized around key theoretical principles

While we are encouraged by our results, we also recognize that learning about the nature of magnetic interactions presents many cognitive challenges. In the case of domains, for example, we found that students associate an element of alignment to a magnetized object, but that their understanding of that alignment might just as likely reside at the atomic level, involve an electrical polarization, or may consist of an alignment which, while similar to the concept of domains, appears more to be an artifact of composition. Students created multiple inscriptions based on the concept of domains, yet appear to not have significantly grasped the concept that each domain is in effect its own little magnet with its own magnetic field.

The nature of nanomagnetic materials poses challenges in learning, both in terms of magnetic fields and domains, but with the added layer of complexity of the particle nature of mater and kinetic molecular theory. We found, for example, that some students believe that atoms in a nanometer size piece of iron are smaller and closer together than in a larger piece. That notwithstanding, the impact of thermal vibrations on the alignment of the atoms in a single (nanoscale) domain was accurately described by a large portion of students. A kinesthetic activity in which students in a group each modeled an atom proved a tangible way to compare surface to volume and the effect of thermal vibrations.

Similar to the findings of Cheng and Brown (2007), we also found that students reverted to prior models in the abeyance of newer more fruitful ones and the tendency to rely too heavily on concrete interpretations of models used for explanatory purposes. Effects of thermal energy or mechanical shock, or dropping a nail to demagnetize it, are often believed by students to physically translate the atoms in the material, rather than to disorient their magnetic alignment. The distinction is confounded, however, with the observation of a magnetic fluid in which the domain particles actually do translate in the absence of a field.

We believe that the study of magnetism provides as meaningful framework and engaging segue for the introduction of nanoscience and of the size dependent nature of materials at the nanoscale. Further study of the ways in which students conceptualize magnetic phenomena across scale, and the cognitive pathways by which those conceptions are able to progress toward greater levels of sophistication will not only provide a descriptive account of learning, but may also inform instruction. In addition, we believe that evidence suggests that magnetism is a meaningful and authentic way in which nanoscience can be introduced to serve existing standards and curricula.

## References

Bar, V., Zinn, B., & Rubin, E. (1997). Children's ideas about action at a distance. *International Journal of Science Education, 19*, 1137-1157.

Barrow, L. H. (1987). *Magnet concepts and elementary students' misconceptions*. Paper presented at the Proceedings of the Second International Seminar– Misconceptions and Educational Strategies in Science and Mathematics, Cornel University, Ithaca.

Borges, A. T., & Gilbert, J. K. (1998). Models of magnetism. *International Journal of Science Education, 20(3)*, 361-378.

Bradamante, F., & Viennot, L. (2007). Mapping gravitational and magnetic fields with children 9-11: Relevance, difficulties and prospects. *International Journal of Science Education, 29(3)*, 349-372.

Catley, K., Lehrer, R, and Reiser, B. (2005). *Tracing a prospective learning Progression for developing understanding of evolution*. Paper commissioned by the National Academies Committee on Test Design for K-12 Science Achievement. Washington D.C.: National Academies Press.

Cavicchi, E. (1997). Experimenting with magnetism: Ways of learning of Joan and Faraday. *American Journal of Physics, 65*, 867-882.

Cheng, M., & Brown, D. (2007, April). *Model construction, critique and revision in magnetism; The importance of integrating conscious and intuitive knowledge*. Paper presented at the American Educational Research Association., Chicago.

Clement, J., & Steinberg, M. (2002). Step-wise evolution of models of electric circuits: A “learning aloud” case study. *Journal of the learning Sciences, 11*, 389-452.

Clement, J. (1989). Learning via model construction and cirticism: Protocol evidence on sources of creativity in Science. In J. Glover, R. Ronning and C. Reynonds (Eds.), *Handbook of creativity: Assessment, theory and research* (pp. 341-381).

Coll, R. K., France, B., & Taylor, I. (2005). The role of models/and analogies in science education: Implications from research. *International Journal of Science Education, 27*, 183-198.

Constantinou, C., Raftopoulos, A., & Spanoudis, G. (2001). Young children’s construction of operational definitions in magnetism: The role of cognitive readiness and scaffolding the learning environment. In J. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*. London: Routledge.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: Sage Publications.

Erikson, G. (1994). Pupils’ understanding of magnetism in a practical assessment context: the relationship between content, process and progression. In P. Fensham, R. Gunstone and R. White (Eds.), *The content of science: a constructivist approach to its teaching and learning* (pp. 80-97). London: Falmer Press.

Gilbert, J. K., & Boulter, C. (1995, April). *Stretching models too far*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco.

Greca, I. M., & Moreira, M. A. (2000). Mental models, conceptual models and modeling. *International Journal of Science Education, 22*, 1–11.

Guisasola, J., Almudi, J., & Zubimendi, J. (2004). Difficulties in learning the introductory magnetic field theory in the first years of university. *Science Education, 88*, 443-464.

Guth, J., & Pegg, J. (1994). First-year tertiary students’ understandings of iron filing patterns around a magnet. *Research in Science Education, 24*, 137-146.

Gutwill, J., Frederiksen, J., & White, B. (1999). Making their own connections: Students’ understanding of multiple models in basic electricity. *Cognition and Instruction, 17*, 249-

Harrison, A. G., & Treagust, D. F. (1996). Secondary students’ mental models of atoms and molecules: implications for teaching chemistry. *Science Education, 80*, 509–534.

Harrison, A., & Treagust, D. (2002). The particulate nature of matter: Challenges in understanding the submicroscopic world. In J. K. Gilbert, O de Jong, R Justi, D. F. Treagust & J. H van Driel (Eds.), *Chemical education: Towards research based practice*. Dordrecht: Kluwer.

Haupt, G. (1952). Concepts of magnetism held by elementary school children *Science Education, 36*, 162-168.

Hickey, R., & Schibeci, R. (1999). The attraction of magnetism. *Physics Education, 34*, 383-388.

Lehrer, R. (2009). Designing to develop disciplinary dispositions: Modeling natural systems. *American Psychologist, 64*, 759-771.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*, 635-679.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. Keith Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 371-387). Cambridge: Cambridge University Press.

Maloney, D. (1985). Charged Poles? *Physics Education, 20*, 310-316.

Nakhleh, M., & Samarapungavan, A. (1999). Elementary school children’s belief about matter *Journal of Research in Science Teaching, 36*, 777-805.

Nussbaum, J., & Novick, S. (1982). Alternative frameworks, conceptual conflict and accommodation: Toward a principled teaching strategy. *Instructional Science, 11*, 183–200.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling and metacognition: making science accessible to all students. *Cognition and Instruction, 16*, 3-118.

Wiggins, G. and McTighe, J. (2006). *Understanding by design.* Alexandria, VA: ASCD.

## Acknowledgments

This work was funded by the National Science Foundation Grant #0426328.

# From Show, To Room, To World: A Cross-Context Investigation of How Children Learn from Media Programming

Thérèse E. Dugan, University of Washington, Seattle, WA, rezza21@u.washington.edu
Reed Stevens, Northwestern University, Evanston, IL, reed.stevens@northwestern.edu
Siri Mehus, University of Washington, Seattle, WA, smehus@u.washington.edu

**Abstract:** We conducted a year-long, naturalistic study that investigated what actually happens when children watch television. We video-recorded children's actions and interactions while watching television and simultaneously recorded the video stream from the television screen; these data were supported with parent diaries and interviews with parents and children. This paper describes two case studies, in which we consider children's interactions with others while watching television and the ways in which their television viewing influences other parts of their everyday lives. We find that both children actively applied knowledge they obtained from visual media to other contexts. In addition, they both shared their media viewing experiences with others, either by directly teaching others about what they had viewed or by creating new content based on what they had viewed.

Television viewing remains an activity that consumes a great deal of young people's time (Kaiser Family Foundation, 2010). Television programming has been criticized for many reasons, including its potential to influence children's behavior and absorb time that might otherwise have been used more productively (Buckingham, 1993; Fisch, 2004; Kirkorian, Wartella, & Anderson, 2008; Maccoby, 1952). Zimmerman, Christakis, and Meltzoff (2007) found that viewing television was associated with slower vocabulary development in children under three years of age. Another recent study (Christakis et al., 2009) found that audible television in a child's environment is associated with lower levels of adult and child interaction. Additionally, Christakis, Zimmerman, Di Giuseppe, and McCarty (2004) suggest that young children's television viewing may have some relationship to attention problems. In response to studies like these, France banned educational television for children under three (France bans broadcast of TV shows for babies, 2008) despite the existence of other research that found benefits of children's viewing (Fisch, 2004; Fisch & Truglio, 2001; Kirkorian et al., 2008; Pecora, Murray, & Wartella, 2006).

This project is informed by studies of how people learn in informal environments and situates itself among similar research studies investigating social interaction and learning (Stevens & Hall, 1998; Stevens, 2000; Stevens, Satwicz, & McCarthy, 2007). Our study directly examines what happens when children watch television in order to understand what types of learning may occur.

In this paper we describe how two children in our study watch and respond to television and YouTube videos, and how their viewing affects other aspects of their everyday lives. We employ an "in show/in room/in world" framework (adapted from Stevens et al., 2007) for the description of children's media viewing and learning. "In show" refers to the television, structure, and conventions of a specific television show. For example, some educational children's shows use conventions, such as dialogue in which TV characters directly address viewers with questions or requests (e.g., to "help" them solve a problem or complete a task), in order to achieve their learning goals. These characters elicit viewer actions by "talking" directly to viewers and leaving time for viewers to respond. Typical prompts include asking viewers to repeat words, actions, and phrases, or to sing along. However, unlike the "in game" designation from Stevens et al. (2007), which was a study of digital games, the viewers of television shows cannot change the action on screen. All interaction occurs entirely in the room. "In room" refers to the social and material characteristics of the environment in which the viewing takes place, particularly the interactions with others that occur there. Finally, "in world" refers to the ways in which the show and viewer responses elicited by the show or by others viewing in the room manifest in other aspects of a child's everyday life.

## Methods

We collected data for 6 months, followed by 6 months of analysis and follow-up interviews. The study included 16 children, 8 boys and 8 girls, aged between 9 months and 6 years. Children were observed and videotaped in their homes for 1-2 hours approximately once each week as they watched different types of media (such as television programming, DVDs, movies, and YouTube videos), and interacted with parents, friends, siblings, toys, and pets. During each session, a videotape record was created and ethnographic field notes were written.

A goal in this study was to record as accurately and completely as possible children's actual experience of watching television in their homes. In order to do this, we scheduled sessions at times when the children were most likely to be watching television in order to avoid interfering with their schedules and routines. We simultaneously recorded the in-show video stream from the television and the in-room activity of children and others in the viewing space (Stevens et al., 2007). We employed a device that synchronized the two streams at the time of recording in the field. In the early stages of the study, we set up the dual-recording equipment and then stepped out of the room while the children were watching. As the study progressed, we were able to further minimize our intrusion by leaving the equipment with the families for a week or more, during which periods the parents turned on the equipment as often as possible when the children watched television. In order to record children when they watched video on a computer, we also used a software program (Screenium) that simultaneously recorded the computer screen and the video stream from the computer's internal camera.

Figure 1: Split-screen image illustrating Harrison "in room" showing his brother Max how to follow the "blast-off" sequence from Leo, and the "in show" *Little Einsteins* program on the television.

We also interviewed or provided written questionnaires to the parents of the children about their child's viewing habits, television learning experiences, and their own feelings about television, movies, and online media usage in their homes. We asked parents to informally interview their children (these interviews were recorded) and we asked specific follow-up questions in a separate questionnaire distributed to parents via email. Finally, parents used journals to record instances of learning and discovery in their children's everyday lives when researchers were not present. These journals were critical with regard to determining whether and how children's viewing habits and interactions when watching visual media affected their everyday activities "in world."

### Data Analysis

All of the video-recorded data were logged and tagged to locate examples of children's responses to television and interactions in the room. Next, we analyzed the actions taking place within each individual clip in greater detail (Stevens et al., 2007), looking specifically at highlighted moments of activity using interaction and conversational analysis techniques (Erickson, 2004; Schegloff, 1998). We devised a transcription method (illustrated in the next section) to distinguish between activities in the show and in the room, and to display activities simultaneously. We used these techniques to illustrate how each viewer action began, what happened during the action, and how long each action lasted. Case studies were created that tied together our video analyses with interviews and diary reports in order to tell the story of an individual child's television viewing experience and relevant learning.

## Case Studies

This section is organized around the case studies of Harrison (age 3 at the beginning of the study and age 4 at the end) and Owen (age 5 at the beginning of the study and age 6 at the end), starting with the television shows they viewed and then examining their in-room interactions and in-world connections to their media viewing behavior. Both children live in the Pacific Northwest, each with a mother and father. Harrison has 1 younger brother and 1 older sister, and attends pre-school every day for a half day. Owen has 1 younger sister and attends kindergarten each day for a full day.

### Harrison

#### Show Viewing Habits

Harrison mainly watched television shows on DVD. In a pre-study questionnaire (April 2008) his mother stated that he usually watched about "*four to six hours of television a week*," including "*Dora and Diego; Super Why? … movies about animals, non-fiction; some Disney movies; basically anything he can get his hands on; no Sesame Street ([he]doesn't like [it]), [and] age-appropriate videos*." Many of the shows he watched were animated shows designed to elicit viewer actions, like *Dora the Explorer*, *Little Einsteins*, and *Go Diego Go*.

In addition to the shows mentioned above that are designed to elicit viewer actions, Harrison also watched non-fiction nature programs such as *Eyewitness Amphibians* and *Henry's Amazing Animals*. These shows employ a narrator and use video footage along with animation to explore animals, their habitats, and their behaviors. In a post-study interview (February 2009), Harrison said he liked "*nature shows the best*" and would watch them over anything else because they "*were real*."

## In Room

Harrison was a very social television viewer in terms of the level of interaction he maintained with other people while watching television shows. He watched shows in his living room, typically with one or more people, who might include his older sister Leanne (aged 5 to 6 during the study), his mother, and his younger brother Max (aged 9 to 21 months during the study). His mother closely monitored what he viewed on television because he did not like to watch shows he had not already seen and because he would get scared easily: "*[he] tends to distrust any show if it has new characters*" (May 2009). Harrison's mother also stated that Harrison usually watched shows that allowed him to learn new facts he could share with her and his other family members. "*[Harrison] enjoys yelling out facts he learns [from shows that] engage him*" (May 2009). When watching television with his sister, his mother said that "*they usually sit by each other and talk about the show. If they get scared, they work together to turn the volume down, hide under a blanket, etc. If they are learning something, they shout out to me. Once in a while they correct each other or bicker a little about a show*" (May 2009).

While viewing shows, Harrison tended to speak to and about the show, regardless of whether the show elicited viewer responses. He responded to shows that prompted him to repeat words, answer questions, repeat actions, and sing songs. Even while viewing shows that lacked prompts for response, Harrison responded to them. Tables 1 and 2 illustrate a sequence of two transcripts, several seconds apart, that exemplify Harrison's response to an episode of a nature show. The top part of the transcript records what is happening sequentially in the show and the bottom of the transcript describes what is happening in the room. The picture in the middle is a frame from the split-screen image from which the segment was taken. Actions on screen are represented in parentheses, while actions in the room are represented in italics.

Table 1: Harrison imitating the movements of the frog's mouth as its froglets jump out (October 30, 2008).

| Participant | 11th sec. | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20th sec. |
|---|---|---|---|---|---|---|---|---|---|---|

**Narrator: are ready to jump out**
**Screen: (*tiny frog jumps out of bigger frog's mouth*)**
***(frog baby lands on the ground)*** ***(slow motion shot of baby frog jumping out of larger frogs mouth)***

*InShow*
*In Room*
**Harrison: (*laying on rocking chair leaning on arm rest gaze towards show*)** **(opens mouth slightly like frog)**

Harrison responds to the nature show both verbally and physically by simulating the movements, actions, and sounds of a frog on the screen. In Table 1, Harrison, while seated on a chair, imitates the mouth movements of a male frog shown on television (which stores its young in its mouth until they jump out fully formed). In Table 2, several moments later, Harrison explains the frog gestation information he just learned to his mother while continuing to watch the show.

Table 2: Harrison explaining the frog's unique gestational practice to his mother (October, 30, 2008).

| Participant | 31th sec. | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40th sec. |
|---|---|---|---|---|---|---|---|---|---|---|

Narrator: After twelve weeks the frog leaves the pond one small hop for the frog one giant leap for evolution
Screen: (*frog leaves pond and hops onto land*) (*frog jumps*)

*InShow*
*In Room*
Harrison: they they grow in its mouth he opens it and then he closes and the frog when its ready jumps out
(*sits up and looks into camera*) (*looks back at show imitates the frog by opening and shutting his mouth*)
(*moves arm to show motion of froglets coming out of mouth*)
(*sits back in chair*)
Mom: wow

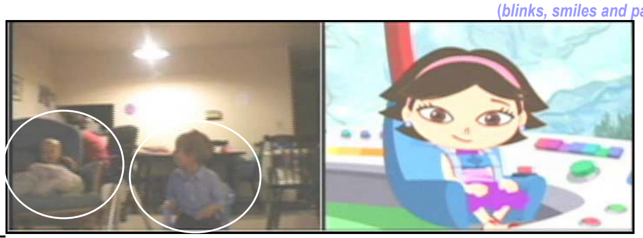
These transcripts reflect Harrison's desire to perform actions along with the show and invite his mother to share in his newfound knowledge about frog gestation. It follows that television shows provide children with information they can learn and teach to others, and that young children who view a television show sometimes imitate and simulate an experience presented in the show, such as how a frog stores its babies in its mouth until they are mature enough to survive on their own. In Harrison's case, he uses his body and voice as he actively responds to the content of the show.

Furthermore, Harrison and his sister are quite familiar with the conventions of certain shows (such as the "blast-off" sequence from shows like *Little Einsteins*), which is the result of "*years of viewing*" according to their mother (March, 2009). Given this familiarity, they are able to coach their younger brother Max (17 months) to share in the experience of responding to prompts from a television show. They demonstrate this by performing the "blast-off" sequence from *Little Einsteins* for Max and he learns to participate appropriately (see Table 3 below). After a few viewings, Max was able to also perform the sequence on his own. Their mother stated that Harrison and Leanne "*encourage him to do actions or sing along*" to shows that elicit viewer responses and in other contexts. They also like to support Max "*…and teach him new words. He's their little prodigy. I think they are excited at the prospect of being able to share their knowledge with him"* (May, 2009).

Table 3: Harrison teaches Max the "blast-off" sequence from *Little Einsteins* (November 25, 2008).

| Participant | 1st sec. | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10th sec. |
|---|---|---|---|---|---|---|---|---|---|---|

Leo: Prepare for blast off we're going to need a lot of power to blastoff pat, pat, pat, pat, pat June is patting,
(*begins patting his lap*) (*continues to pat lap*)
Rocket: (*Rocket wiggles*)
June: (*blinks, smiles and pats lap*)



*InShow*
*In Room*
Leanne: Max remember? pat, pat, pat , pat, pat, pat, pat Max likes it
(*leans forward and stretches out arm and sits up on knees and begins patting her lap*)
Harrison: pat, pat, pat hee hee hee
(*camera pans to Harrison who is patting and looking at Max*)
Max: pa
(*camera pans to Max patting lap*)

In sum, Harrison was an active viewer who responded verbally and physically to television shows that were designed to elicit responses, and to shows that were not designed to elicit any responses. According to his mother's journal, as well as our interviews and observations with him and his family, as Harrison viewed specific television programs designed to elicit viewer responses he developed techniques for how to act while watching those programs. He passed along these techniques by encouraging his younger brother to also share in the experience of responding to a show that elicited viewer responses. He also liked to share facts and information he learned from watching nature shows with his family, and applied these skills in his interactions with his family while watching television and in other contexts.

### In World

During the study, Harrison also began applying information learned while watching one show to other experiences, including other shows that did not elicit viewer responses. For example, his family went on a trip to Arizona to visit his grandparents. While there, they all viewed *Disney's Planet Earth* as a family. Harrison's mother said she was stunned when Harrison began to name the locations of certain waterfalls and demonstrate knowledge about many of the caves, animals, and rock formations in the show. She stated that Harrison and Leanne said they "*learned*" about them from *Dora the Explorer*, *Go Diego Go*, and *Little Einsteins* (March 2009). "*This instance was unusual because it crossed the bridge between cartoons and non-fiction videos,*" She thought it was interesting that Harrison was discussing information he saw and learned about in an animated fictional show and was relating that information while watching a documentary about nature (May 2009).

In addition, Harrison's mother stated that television viewing helped him make learning connections away from the television set. In her viewing journal and interview, Harrison's mother stated that "*about twice a month [Harrison] will randomly say something that he 'learned' from TV.*" For example, "*if we go to the zoo, he will say facts about the animals*" (May 2009). Furthermore, in her parent journal she wrote that Harrison asked to watch shows such as *Reading Rainbow* because *"they helped him learn to read*" (December 2008). She also said in a post-study interview, "*we did get some Diego phonics readers…because he asked for them"* (May 2009). She believed that, when Harrison asked for a Diego book for Christmas, he thought it would help him learn to read. So Harrison connected the shows he viewed to other aspects of his life.

## Owen

### Show Viewing Habits

Owen, who turned 6 years old during the study, did not watch any shows that elicited viewer responses in our study observations, yet he frequently responded to the shows he watched. His mother indicated in her journal and interviews that Owen previously watched shows such as *SuperWhy* and *Little Einsteins* that elicited viewer responses when he was a bit younger (April 2008). In a pre-study questionnaire (April 2008) his mother stated that Owen usually watched about "*fifteen to sixteen*" hours of media each week, including television shows on PBS such as "*Curious George; Dragon Tales; whatever is on when he comes home*," as well as videos on YouTube about "*Bionicles movies; commercials; vintage tv ads for toys; reviews of toys (home movies), for example Bionicles (showing how to put toys together)*," and movies such as "*…Pokemon, Digimon movies; The Black Stallion; [and the] Marx Brothers*." We observed Owen mainly watching shows on the afternoon block of PBS Kids (*Curious George*, *Arthur*, *Dragon Tails*, *Word Girl*, and *Fetch with Ruff Ruffman)*; we also observed him watching toy reviews on YouTube for Lego Bionicles as well as Hot Wheels cars and Transformers. Furthermore, we observed Owen viewing the live-action nature documentary series *Nature: The Beauty of Ugly* and the animated DVD series *Avatar the Airbender.*

In a post-study interview (May 2009), Owen stated that his favorite shows were "*Curious George because he can do things we can't*," "*Fetch with Ruffman because it's funny,*" and "*…Nature…because…my favorite thing [is] the rat-tailed scorpion… they, they were – they're not…insects like scorpions, but they're not, they're not like spiders but they're a different species… and they look weird.*" The nature shows were generally more realistic than the other types of programming that Owen viewed in terms of their settings (real life vs. cartoon fantasies).

### In Room

On an average day, Owen watched media for up to 4 hours in his basement. He was usually unaccompanied when he watched video, although he sometimes watched shows with a friend, with his mother and younger sister Mia (who was 1 year old) in the afternoon, or with his father before bedtime.

Owen often displayed sensitivity to what he saw on the screen. On several occasions Owen became upset when he viewed show characters in embarrassing or suspenseful situations. His responses to these kinds of situations included running away screaming, hiding his head under a pillow or in his hands, or shouting that he hated it or that it made him scared. With regard to these observations, Owen's mother stated that "*he seems to be frightened when something 'bad' is about to happen, whether that be someone getting angry or someone getting hurt. It does confuse me. The best answer I can give right now would be that he takes things very literally. He doesn't do a lot of 'pretending to be a character'. He gets frustrated and confused when his friends claim something that is just not true. Also, he has an explosive personality himself, and so maybe he is afraid that the characters in the show would react the same way he would, and that scares him. [Owen] feels things very deeply, so if there are scary things happening he will feel very scared!*" (May 2009).

Owen developed coping mechanisms that he employed as he watched television, even as he watched programs that were designed for children his age and that did not contain purposefully frightening content. Table 4 illustrates how Owen coped with a segment from *Curious George* that scared him.

Table 4: Owen copes with an unsettling situation in Curious George (January 12, 2009).

| Participants | 11th sec. | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20th sec. |
|---|---|---|---|---|---|---|---|---|---|---|

1. Narrator: George didn't want to get in trouble but he
2. George: (looks up) (stands with hands clutched)
3. MWYH: just tell the truth no one will be upset
4. Dog: (ooks displeased and frowns)

5.
6. *InShow*
7. *In Room*
8. Owen: tiss huhh I can just see it I can't do this ok I'll look again
9. (covers eyes with hands) (kicks up leg and lunges towards couch)
10. (sighs; uncovers eyes)

While watching *All New Hundley*, an episode of *Curious George*, in which Hundley the dog messes up the room and George is accused of making the mess, Owen was visibly upset by the fact that George was accused of something he did not do. Owen got through the unsetting situation by telling himself (out loud) that he could not look, but then decided to make himself start viewing again to see whether George would be exonerated (line 8). This behavior matches his mother's observation that Owen disliked scary and misleading situations.

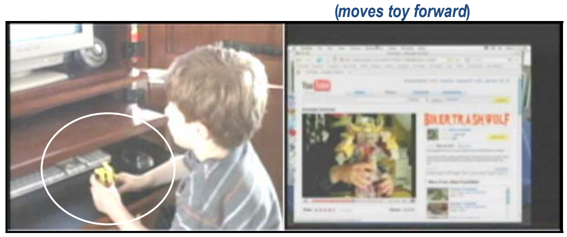
According to the PBS Parents Web site, the educational objective for this episode was to "demonstrate inquiry about the world and solving problems," a process that "includes a variety of skills," and to "model positive attitudes toward learning like persistence and curiosity" (All New Hundley). Owen's reaction suggests that he missed the show's intended message because of the context in which that message was presented. Rather than reflecting an interest in persistence and curiosity, Owen relied on his technique for coping with the emotionally challenging content of the show, and engaged with the show again after George had been cleared of wrongdoing. Therefore, the way in which Owen shared George's experience was, most likely, not what the producers of the show intended.

Other forms of media also have a powerful impact on Owen, such as the toy review videos that he watched on YouTube. Owen viewed toy reviews because he found them interesting and wanted to learn about specific toys and their differences to aid his decision-making process for purchasing new toys. As Table 5 illustrates, Owen used the computer in his living room to watch a YouTube toy review about a Transformer.

Table 5: Owen manipulates Transformer toy when watching YouTube toy review (July 23, 2008).

| Participants | 1st sec. | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10th sec. |
|---|---|---|---|---|---|---|---|---|---|---|

1. Biker T: And that's him Rotate the other part of this plan it have right here?
2. Turn them around to expose the orange cuz I like the orange better than the purple
3. (touches head of toy) (adjusts plate on transformer leg) (adjusts plate on other leg)
4. (moves toy forward)



5.
6. *InShow*
7. *In Room*
8. Owen: (holding transformer and watching video) (turns his transformer to the right) (moves transformer back to face video)

As the video narrator discussed the toy he was reviewing, Owen (line 8) manipulated his own similar but smaller toy along with the narrator's description. The narrator did not ask viewers to have a similar toy handy during his review, yet Owen had his toy with him anyway. Owen said he always tried to have a toy or object similar to the one being reviewed so he could more carefully inspect his own toys and see how they compared to those being reviewed (May 2009). He told us he considered these reviews useful because '*you know what to expect*' before buying something (May 2009), which could indicate that Owen was making a judgment about whether a given toy was different enough from his current toys to warrant a purchase.

The notable difference between Owen's behavior in these two examples is the level of emotional arousal he exhibits while viewing *Curious George* and the toy review. The *Curious George* show elicited a highly-charged emotional, empathetic reaction that seemed quite visceral and difficult for Owen to control. He reacted in a way that, according to his mother, mirrored several other outbursts, many of which resulted in a temporary cessation of viewing. By contrast, his viewing of the YouTube toy review was steady and controlled from an emotional perspective, and Owen's activity was limited to his manipulation of a similar toy. Instead of responding reactively, Owen ultimately responded proactively to the toy reviews he watched.

### In World

As Owen viewed more toy reviews on YouTube, he engaged directly with this form of media by becoming a toy reviewer. He came to the conclusion that other toy reviewers did not always provide helpful reviews, and he began to examine more closely how people reviewed products and how good and bad toys were differentiated. After a while, according to his mother, Owen wanted to let people know what he learned, so he recorded his own review of Jazz, a Transformers toy that he enjoyed playing with, and posted his review on YouTube. He stated that he wanted to create a review that would give viewers a more honest assessment of the product they were researching compared to some of the other reviews (May 2009). After posting the video, a few people left complimentary and encouraging comments for Owen on his YouTube page that stated how helpful his review had been for them (see Figure 2).

Figure 2: Owen's YouTube Toy Review

In this context, Owen developed proactive viewing skills that were not reflected in his highly reactive style of viewing television shows such as *Curious George*. Owen learned to watch toy reviews with a critical eye and he chose to apply that learning in practice by becoming a producer of reviews for others to watch. As stated earlier, television shows and YouTube videos provide children with information they can learn and, subsequently, teach to others. In Owen's case, the YouTube toy reviews that he watched filled in gaps in his desired knowledge about a toy that he owned, gaps that neither his parents nor his teachers could fill. His viewing of these videos also triggered an interest in producing a video review of his own with distinctive features, a video that generated enough interest to accumulate over 1,000 views and a 3-star rating.

## Discussion and Implications

While current research (Christakis et al., 2004, 2009; Zimmerman et al., 2007) indicates that young children are passive viewers who should not watch television because of potential language development and attention problems, the present study provides a more nuanced perspective. This study is important in part because little is known about how young children watch television in natural settings. Our observations demonstrate that young children can be active viewers who are indeed learning from their experiences with media and who act to draw others into their viewing experiences. Our study suggests that we might want to worry less about the mere fact of whether children watch television but rather pay more attention to what they do with it and how that might be cultivated in ways that support their development as people (cf. Penuel et al., 2009).

A good assessment of learning is the ability to actively apply knowledge from one context to another. Regardless of the media formats that Harrison and Owen viewed, both boys actively applied knowledge they obtained from those media to other contexts. The similarities between these boys center around their sharing of media viewing experiences with others, either by helping others understand television shows they viewed or creating new content they considered superior to what they had watched.

Harrison and Owen both developed specific, active viewing habits while watching and responding to television shows. Furthermore, both boys seemed to engage emotionally through embodied responses when viewing these shows. After repeated viewing of children's television shows that elicited responses from young viewers, Harrison developed the skills needed to interact with television shows and he even taught his younger brother the same skills. Moreover, when viewing a show Harrison imitated and acted out the movements and actions of the animals on screen. Harrison also shared the facts he learned from shows with other members of his family, a desire to teach that resembles Owen's proactive creation of YouTube toy reviews. Owen was also an active and sensitive viewer who developed ways of exerting control in relation to visual media. He taught himself coping techniques for viewing some television content that he found disturbing, and he recorded and published on YouTube his own toy review.

## References


All New Hundley (n.d.). Retrieved October 13, 2009 from http://www.pbs.org/parents/curiousgeorge/program/ep_desc_s2.html.

Buckingham, D. (1993). *Children talking television: The making of television literacy.* London: Falmer Press.

Christakis, D. A., Zimmerman, F. J., Di Giuseppe, M., & McCarty, C. A. (2004). Early television exposure and subsequent attention problems in children. *Pediatrics, 113*(4), 708-713.

Christakis, D. A., Gilkerson, J., Richards, J.A., Zimmerman, F. J., Garrison, M. M., Xu, D., Gray, S., & Yapanel, U. (2009). Audible television and decreased adult words, infant vocalizations, and conversational turns: A population-based study. *Archive of Pediatric and Adolescent Medicine*, *163*, 554-558.

Erickson, F. (2004). *Talk and social theory: Ecologies of speaking and listening in everyday life.* Malden, MA: Polity Press.

Fisch, S.M. (2004). *Children's learning from educational television: Sesame Street and beyond*. LEA's communication series. Mahwah, N.J.: L. Erlbaum Associates.

Fisch, S.M., & Truglio, R.T. (2001), *"G" is for growing: thirty years of research on children and Sesame Street*. Mahwah, N.J.: Erlbaum.

France bans broadcast of TV shows for babies (August 20, 2008). Retrieved October 13, 2009 from http://www.msnbc.msn.com/id/26312386/.

Kaiser Family Foundation. (2010). *Generation M2: Media in the lives of 8- to 18-year olds*. Retrieved February 27, 2010, from http://www.kff.org/entmedia/upload/8010.pdf.

Kirkorian, H.L., Wartella, E.A., & Anderson, D.R. (2008). Media and young children's learning. *Future of Children, 18*(1), 39-61.

Maccoby, E.E. (1952). *Television: its impact on school children.* Indianapolis: Bobbs-Merrill.

Pecora, N.O., Murray, J.P., & Wartella, E. (2006). *Children and television: 50 years of research*. Mahwah, N.J.: Lawrence Erlbaum.

Penuel, W. R., Pasnik, S., Bates, L. Townsend, E., Gallagher, L. P., Llorente, C., & Hupert, N. (2009). *Preschool teachers can use a media-rich curriculum to prepare low-income children for school success: Results of a randomized controlled trial.* Newton, MA: Education Development Center and SRI International.

Schegloff, E. A. (1998). Reflections on studying prosody in talk in interaction. *Language and Speech, 41*, 235-263.

Stevens, R. & Hall, R. (1998). Disciplined perception: Learning to see in technoscience, *Talking mathematics in school: Studies of teaching and learning*, pp. 107-149. M. Lampert & M. L. Blunk, (Eds.), Cambridge University Press: Cambridge.

Stevens, R. (2000). Divisions of labor in school and in the workplace: Comparing computer and paper-supported activities across settings. *Journal of the Learning Sciences, 9*(4), 373-401.

Stevens, R., Satwicz, T., & McCarthy, L. (2007). In game, in room, in world: Reconnecting video game play to the rest of kids' lives. In Salen, K. (Ed.), *Ecology of games: MacArthur Foundation series on digital media and learning.* Cambridge, MA: The MIT Press.

Zimmerman, F.J., Christakis, D.A., & Meltzoff, A.N. (2007). Associations between media viewing and language development in children under age 2 years. *Journal of Pediatrics*, *151*, 364-368.


## Acknowledgments


This work was supported by a National Science Foundation grant (NSF#0354453). Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation. For more information on the LIFE Science of Learning Center see: http://life-slc.org/. Furthermore, Stephanie Scopelitis, Sheldon Levias, and Lionel Enns contributed to data collection for this study. Finally, the authors would also like to thank Professor Stevens' Learning, Media, and Interactivity (LMI) Research Group for their helpful comments and ideas.

# Arts and Learning: A Review of the Impact of Arts and Aesthetics on Learning and Opportunities for Further Research

Kylie A. Peppler, Heidi J. Davis, Indiana University, 201 N Rose Avenue, Bloomington, IN, 47405
Email: kpeppler@indiana.edu, hdavis2@indiana.edu

**Abstract:** Traditionally, learning scientists have paid little attention to the discipline of the arts as the more prominent focus has been on science and mathematics. Despite this, the learning sciences could benefit from further understanding how the arts offer alternative methods of inquiry, representation, and understanding. At the same time, leaders in the field of arts education are calling for more research in areas that intersect with the learning sciences, pointing to the mutually synergistic ways that the two fields could inform one another. Guided by feminist communitarian methodology, this paper brings together a review of a wide body of research in the field of arts education, spanning affective, cognitive, social, and transformative effects of the arts. Insights are shared for how the arts as a discipline can inform the study of learning and, conversely, point to ways in which learning scientists can contribute to the field of arts education research.

The arts and aesthetic traditions have long since been a crucial part of the education experience. Scholars, such as John Dewey, have written extensively on the role of the arts in education (1934/1980), conceiving of the art object, the process of engaging in art making, and the resulting aesthetic sensory experience as theoretically independent entities. According to Dewey, the activity of art making is important because it engages learners in the process of building, designing, and constructing artifacts and provides a tool by which we search for meaning. Both of these notions lie at the heart of a great deal of learning sciences research. Despite this history, learning scientists have paid little attention to the discipline of the arts as the more prominent focus has been on science and mathematics. There are few art educators found in the learning sciences literatures and there are few scholars identified with the learning sciences community who deal with the arts in an explicit way. Despite this, there are many aspects of learning that could benefit from further research into how the arts and aesthetics play a role in our current conceptions, including the study of representations (Latour, 1987; Lynch, 1988), the design of new experiences, environments and technologies (Brown, 1992), and visual research methodologies (Pink, 2001), among many other areas of learning sciences research. One example that seems particularly ripe for exploration is the study of representations, particularly how children engage in drawing, dramatic re-enactments, and role-play when representing their ideas in simulations or models (c.f., Danish, 2009). Artistic and aesthetic facets are innate in all of this work, while their role in the learning process is seldom explored.

Historically, cognitive views of learning have prevailed in arts education (Deasy, 2002; Gardner, 1991) but more recently views have expanded to take on increasingly diverse theoretical frames and areas of study related to learning (Alexander & Day, 1992; Eisner, 2002). Taken together, these bodies of research have contributed to our understanding of the role of the arts in (1) the methods of data collection and analytical techniques (Barone, 2008; Sullivan, 2005); (2) the healing, therapeutic, and restorative processes (Eisner, 2002; Malchiodi, 1999; McNiff, 2004); (3) the discovery of the self, the group, the community or diverse cultures (Greene, 1995; Dewey, 1934/1980; Heath & Soep, 1998); and (4) in expression and communication as well as the particular habits of mind that are cultivated through the arts (Hetland et. al, 2007). More recent studies point to the long-term impacts of the arts on learning and development that outweigh other extracurricular activities (Catterall, 2009). These bodies of research are summarized more fully in this article but, taken together, represent a robust body of knowledge that is rarely built upon in learning sciences research. At the same time, leaders in the field of arts education are calling for more research on learning and artistic expression as well as more research on the arts and the transfer of learning (Arts Education Partnership, 2004). This points to the mutually synergistic ways that the two fields could inform one another. With this community bond in mind, we adopt the feminist communitarian model as our guiding methodological framework. This framework celebrates community ties and conceives understanding as inseparable from community (Friedman, 1989).

The focus of this paper is to turn our attention to the arts as an understudied area and examine how the field can expand as well as open new avenues of research in the learning sciences. This paper brings together a review of a wide body of research in the field of arts education, spanning affective, cognitive, social, and transformative effects of the arts. In addition, insights are shared for how the arts as a discipline can inform the study of learning in educational settings and, conversely, point to ways in which learning scientists can contribute to the field of arts education research.

## History of Arts Education

The study of the arts in education has a long history arcing back to the Greek and Roman eras. Plato and Aristotle wrote of the importance of the arts, "for their didactic impact as instruments of cultural maintenance," (Efland, 1990, p. 8), alluding to the earliest function of the arts as visual public curricula to convey cultural values and encourage public discourse. In his influential book, *A History of Art Education*, Efland (1990) depicts the ways in which the social status of the artist contributed significantly to how art education has been conceptualized throughout time. For example, while art objects were seen as important for the education of Greek and Roman societies, the rich did not participate in art making, for artists were not deemed important in society. Conceptions changed in the Middle Ages, when learning in the arts became a form of religious penance for many monks, thus shifting the status of learning in the arts from low-status to highly esteemed. The next big shift occurred during the Renaissance, when academies were created to educate learners on the philosophy of art, which separated art making, traditionally done by lower status individuals, from theorizing about art, now part of the newly formed academies. As these two learning contexts developed, separate identities emerged: craft and fine art. No longer appropriate only for the working class and clergy, the concept of fine art brought high status to learning in the arts and opened the door to public art education.

In the United States, art education began with emphasis on drawing in Benjamin Franklin's academy in Philadelphia (1739/1931). The Industrial Revolution also encouraged drawing as a technical skill that should be included in the general education setting. Similarly, art at the university level began in other departments such as science and anthropology. For example, drawing was a subject taught in the School of Science at Yale long before the School of Fine Arts was established. At the turn of the twentieth-century, as fine art schools emerged across the country, technical skill and art history dominated learning in the arts. At the same time, Dewey's work on aesthetics entered the educational discourse, linking all learning to aesthetic experiences and aesthetic experiences to art in his seminal work, *Art as Experience* (1934/1980). As aesthetic awareness entered the stage of learning in the arts, somatic and intuitive understanding began to get the attention of much research. Research on the imagination and creative expression emerged from this pool of research, as well.

In more recent years, these discussions have been eclipsed in some of the current day educational policy imperatives that place central the role of science, math, and literacy in 21$^{st}$ century education. Often missing from these conversations is the enduring role that the arts and aesthetic knowing plays in education. Aesthetic awareness concerns the heightening of sensory perception (Greene, 1995), enabling learners to discern and demarcate something temporal as rare, unexpected, or beautiful. Arts researchers often point out that art-centered experiences are not automatically aesthetic in nature, but rather meaning is developed through the reflective and critical eye (Dewey, 1934; Freedman, 2003). Desantis and Housen (2000) developed a theory of aesthetic development that outlines patterns of thinking that correlate with the amount of art-centered experiences a learner has had. The theory presents a continuum in which lower levels of aesthetic connection align with fewer exposures to art. This work points to the importance of art education for aesthetic growth, which is of central importance to studies of arts learning, as well as in other domains.

## Methodology

What follows is a brief review of the vast body of research that has taken place in the field of arts education. At its outset, four guiding questions framed our review and helped to organize our writing, consisting of: (1) What counts as learning in the arts? (2) How has learning been historically studied in the arts? (3) As an understudied area in the learning sciences, what implications do the arts hold for future research? By contrast, (4) how can the field of learning sciences expand the efforts of arts education?

Our intent for this review is to precipitate dialogue that bridges the educational discourse in two communities of research. In that regard, we selected a methodology anchored in communitarian epistemology apropos for building community ties and mutual interest. Our research methodology emanates from the feminist communitarian model, which puts forward that "the community is ontologically and axiologically prior to persons" (Christians, 2005, p. 152). This model provides three key foundations for this paper: where to look, what is important to look at, and how to use this information. First, this model provides a sociocultural lens, which suggests that meaning is mediated through social dialogue. From this perspective, we began the literature review by looking to key portals of dialogue in the arts community. As such, we bring together several landmark studies found in books and major journals published by leading arts education organizations, including the National Arts Education Association (NAEA) and the Arts Education Partnership (AEP).

Second, our methodological framework guided the process of selecting key themes for the paper. The feminist communitarian model holds that rules situated within a community are understood by valuing the multiplicity of voices as opposed to formal consensus. Within this framework, we looked to multiple voices in the social dialogue that provided themes that are unique to the arts. Taken together, this body of research can be summarized in four unique properties of arts learning that have broader implications for learning more generally, which are summarized in Table 1. In the following sections of the paper, we go into each of these four themes in further depth, highlighting some prior landmark studies as well as pointing to ways that the work can inform our understanding in the learning sciences.

## Art as Inquiry

**Arts-based researchers** look to aesthetic response within the disciplines of art in order to develop meaning. Arts-based research is a methodology, which includes art making into any of the phases of research from inquiry and data collection to sharing findings and conclusions (Deasy, 2009). Poetry, narrative, film, drawing, collage, painting, performance, dance, music, and sculpture are a few of the mediums used in the creative research process. Arts-based research can be used to explore questions in any discipline, not just inquiries about art. Arts-based research critically investigates learning based on the belief that there are multiple ways to develop meaning. Here, we focus on arts-based research in education, though it exists in other disciplines, such as engineering (Penny, 2000), and anthropology (Pink, 2001).

Table 1: Summary of the current research organized into four key themes.

| Properties of Learning in the Arts | Definition | Implications for the Learning Sciences |
|---|---|---|
| Art as Inquiry | The methods and mediums used by artists offer tools for art as a way of research. | Qualitative research methods. |
| Art as Transformative Experience | Art-centered experiences offer therapeutic and restorative properties. | Affective understanding.<br>Self-efficacy.<br>Learner readiness.<br>Restorative experience.<br>Healing. |
| Art as a Means of Discovery | The arts provide opportunities for multiple solutions, which allow individual differences to emerge and allow individuals to be reflexive about the self, the group, the community, and diverse cultures located across space and time. | Community Building.<br>Collaboration.<br>Cross-Cultural Awareness.<br>Sociocultural Theory. |
| Art as Literacy | Engaging in the arts requires specific habits of mind and knowledge-in-action to be able to read and write the arts. | Habits of Mind.<br>Literacy.<br>Critique.<br>Representations.<br>Transfer. |

There are many definitions and labels for inquiring about the world through artistic practice. Most descriptions include the notion of viewing education through the artistic lens. But that means slightly different things to different researchers. Sullivan (2005) summarizes four ways that education research has been interpreted through the arts: 1) educational connoisseurship, 2) Arts-based Educational Research, 3) arts-informed research, and 4) a/r/tography. A fifth group is added to Sullivan's list by including the scholARTists mentioned in Cahnmann-Taylor and Siegesmund's collection of arts based research (2008).

Sullivan's first group, Eisner's educational connoisseurship, is based on *seeing* and *sensing* educational phenomenon through the expertise of an *educational critic*, modeled upon the concept of the expertise of an art critic, who makes judgments based on having exposure to many images Similarly the educational critic makes judgments based on exposure to many educational experiences. Eisner continued to push the idea of art practice informing research by creating the Arts-Based Research Institute in 1993. In 1997, with his colleague Barone, the concept and the term, Art-based Educational Research (ABER), emerged. ABER points to the unique insights and ways of expressing that can be found only in the arts..

While ABER celebrates the distinctly different approaches available through the arts, other researchers maintain the traditional research framework for using the arts in research. For example, arts-informed research seeks to integrate scientific inquiry with artful and imaginative inquiry (Cole & Knowles, 2001). From this view, the arts are relied upon to represent understanding developed through more traditional research methods. To distinguish between ABER and arts-informed research, the results of an arts-informed study might be generalizable, while ABER would not seek generalizability, but instead value *aesthetic resonance* --that is, the research is recognized as holding meaning for a specific community (Eisner, 2008). Other researchers refer to themselves as a/r/tographers, who inquire about the world within the identity of being an artist who is also a researcher and a teacher. A/r/tographers approach research through living inquiry embodied within the daily roles of an artist-researcher-teacher. As such, a/r/tographers use multiple artforms and writing to build understanding by knowing, doing, and making (Irwin, 2004). In addition to these four interpretations of art as inquiry depicted by Sullivan, another key group includes ScholARTists, practicing artists who create stand-alone artwork developed from their research. Their artwork develops from fieldwork, such as collected and analyzed data, but the final piece is relayed through art. For example, Saldana (2008) created an ethnodrama about the nature of qualitative data based on field notes and 25 years worth of experience in the theater. ScholARTists use art practices for scholarships' sake (Irwin, 2004).

All arts-based research aims to establish research questions, share findings, and mine data in evocative and meaningful ways (Cahnmann-Taylor & Siegesmund, 2008; Irwin & de Cosson, 2004). For example, Springgay's (2004) dissertation research on tactile epistemologies employs the methods of a/r/tography to examine student perceptions of *thinking through the body*. The paper version of the dissertation is accompanied by a DVD of seven videos. Aligning with the tactile epistemological content of the work, the paper version is hand stitched with fabrics and thread, and is augmented by the author's original photography and poetry.

Ultimately, research is a tool for making sense of learning in order to positively impact the lives of learners. Human understanding advances through exploration, interpretation, and representation, which are qualities of both the arts and sciences (Leavy, 2009). Arts-based research is beneficial for the learning sciences in myriad ways, but we have highlighted three of the most significant: 1) to provide the learning sciences with an expanded toolkit for developing meaning, 2) to acknowledge and clarify assumptions about what it means to know for the researcher, and 3) to include voices previously left out of the discourse on learning. In addition to the benefits of embracing aesthetic awareness as a useful component of research, arts based research also challenges other traditional notions. Embracing post-modern theories that state there is no one truth, arts-based research encourages research that opens the conversation--it is not the researcher telling the reader, but rather the researcher invites the reader to be an active part of the process in making meaning (Siegesmund & Cahnmann-Tayor, 2008). Barone also sees the wider readership potentially available through arts-based research as an opportunity to encourage social change (2008).

## Art as Transformative Experience

Throughout history, humanity has responded to tragedy with art, as represented by the 9/11 memorial, *Reflecting Absence* or Picasso's Guernica, among countless others. Just as a nation uses art to heal and move forward, art as a transformative experience is important for education, as well. As Maslow's hierarchy of needs suggests, basic needs such as feeling safe and valued must be met before learning can occur (1943). The transformative nature of learning in the arts is well poised to meet this educative need.

Long acknowledged in the fields of art, psychology, and aesthetic education, the process of creating and responding to art offers therapeutic and restorative properties (e.g., Eisner, 2002). The transformative impact of art can be seen in studies like Malchioldi's (1999) work with children, which found stress relief and healing effects when art-making was used with hospitalized children. Kaplan et al. (1993) found that visits to a museum had a restorative effect on visitors. A number of self-proclaimed transformative art experiences exist in the literature, as well. For example, Hill (1951) professes that art making was instrumental in his recovery from tuberculosis and Allen (1995) shares how art helped her confront harmful beliefs and transformed her life.

The transformative impact of art provides insight for research in the areas of affective knowledge, restorative experiences, and self-efficacy. Art therapy is a branch of psychology that has greatly contributed to this range of research. Art therapy research has shown that healing and transformation comes from a place of affect that is beyond words (Malchioldi, 1999). Research has also shown that the affective benefits of learning in the arts act as a gatekeeper for positive self-identity in adolescents (Catterall, 2004). Similarly, in *Critical Links*, several research projects found that the affective properties of the arts contributed to positive transformations in the areas of motivation, cognitive development, school culture, academic performance, and attitudes toward reading (Catterall, 2002). Self-efficacy is also strengthened through experiences with art (Catterall, 2004; Catterall & Peppler, 2007). Eisner summarizes that self-efficacy is developed during the creative act as learners engage in a medium and bring their concepts into reality (2002). Kennedy's (2002) study found that learning in the arts increased the self-efficacy of at-risk youth, promoting positive self-esteem. In this way, experiences with art enhance individual, social, and cultural health. By readying the psychological state, learning in the arts reaches across disciplines and offers unique opportunities to restore health, which clears pathways to new knowledge.

## Arts as a Means of Discovery

Another key contribution of the arts is that it provides opportunities for multiple solutions. As Eisner states, "…Standardization of solution and uniformity of response is no virtue in the arts. While the teacher of spelling is not particularly interested in promoting the student's ingenuity, the art teacher seeks it" (2002, p. 1). The diversity of solutions and the space that is afforded for creative solutions to a problem is at the heart of what it means to engage in learning in the arts. We see the valuing of multiple solutions as essential to allowing individual differences to emerge and engender a discovery learning process that engages the learner in learning about the self, the group, the community, and diverse cultures, which are outlined in the four sections below.

*Arts and Learning about the Self.* While self-discovery may happen in all disciplines, the arts seem to be a particularly fruitful context to cultivate solutions unique to the self, imagination, and creativity. Further discovering these areas in terms of their relationship to learning seems ripe for exploration into the role of creativity and imagination in learning and seems to also be an understudied area of the learning sciences. Imagination seems to play a role in self-discovery. Greene's seminal work, *Releasing the Imagination*, stresses that when a young person's imagination is not released, that young person may have difficulty situating the self as well as the role of the self in a larger community (1995). Heath and Soep expand on this to elaborate that the arts allows individuals to be reflexive about the self as they hone the ability to make things of value to their surrounding communities (Heath & Soep, 1998). Similarly, Dewey alerts us to the transformative nature of the arts and aesthetics in challenging the status quo and the dominant elite in order to meet the needs of democratic society (1934/1980). This is particularly relevant to youth in marginalized communities because they have an opportunity to write their own narratives and insert themselves into the dominant discourse through the arts. This potentially sets the stage for higher levels of engagement in other arenas, like school. Catterall and Peppler also discuss the impact that the arts have on general self-efficacy in disadvantaged groups—the positive and authentic view of one's capabilities and achievements—developed in mastering an art form, and the critical and reflective dispositions that accompany its development (2007). Taken together, these strands of research call for future investigations into the differing effects of an arts experience for the audience and the artist, which are areas of inquiry that learning scientists are particularly poised to answer.

*The Arts and Group Learning.* As aforementioned, when young people engage in art making they explore a variety of disciplines and are learning at multiple levels, including learning about the larger group or classroom community as well as their place in it. Arts experiences frequently involve more than one learner. The performing arts, in particular, are steeped in this tradition as actors find their place in the production, musicians learn about their part in the orchestral work, visual artists work on large murals together, and dancers in their role in the dance. While the arts are not the only discipline to have group learning experiences, they offer rich opportunities to gain skills in a group setting and also to display final products. Research on collaboration would benefit from exploring the arts further to better understand the role of the collective in the arts. Researchers might also inquire into the specific qualities the arts bring to group learning processes as well as the qualities of the individual art forms that contribute to efficacious group learning and contrast this with what is known more generally about group learning and collaboration. Other areas that are ripe for exploration include investigating whether arts collaborations exhibit general tendencies to enhance equity in learning for larger numbers of learners in the group and particularly whether this is inclusive of already disenfranchised groups. Preliminary observations indicate that the arts can create more equitable learning opportunities for at-risk youth (Catterall & Peppler, 2007). Additionally, it would be interesting to explore whether the arts can serve as a training ground for learning to be part of a group outside of the arts and the conditions for such group learning to occur.

In a related manner, current and historical research points to the arts as the foundations of a democratic society, including effects of the arts on positive social interactions, tolerance, and consideration to moral dilemmas. For instance, studies suggest that the arts promote empathy, tolerance, and solution finding through taking multiple perspectives (Catterall, 2002). These effects may not just extend to students involved in the arts, they may well impact participating teachers and school identity. For example, Noblit and Corbett (2001), noted in their evaluation of the A+ Schools program in North Carolina that school faculty developed a positive school culture despite typical administrative challenges and lack of resources. This work suggests that engagement in arts activities fosters democratic values. Further inquiry may advance notions of democratic public schooling.

*The Arts and Cross-Cultural Learning.* Because artifacts are a reflection of the values held by a group of people in a particular space and time, they allow us to learn about diverse cultures through their study. This type of learning happens even when we travel across space and time. Anthropologists and historians, for example, help us to understand diverse cultures through the study of their art, which is the foundation of fields like art history and music history. Moreover, the arts are rooted in cultural traditions. Engaging youth to identify more deeply with their own culture and share this understanding with others may also be a key contribution of the arts, especially as classrooms encourage this type of sharing with peers and teaching faculty. Moreover, youth from diverse backgrounds can develop deeper understandings around issues of race, culture, and class systems (Deasy, 2002). Drama, for example, has been found to engage youth in social change and build understanding among diverse groups (Rohd, 1998). This may be because drama allows youth to explore multiple roles and perspectives through role-play (Deasy, 2002). In doing so, drama helps youth to understand character motivation, complex problems and emotions, and social relationships, promotes conflict resolution, engagement, and positive self-concept (Catterall, 2002). This ongoing body of research demonstrates the efficacy of the arts of communicating meaning across cultures through complex semiotic systems.

## Arts as Literacy

Literacy is now known to be both multimodal in nature (Kress & van Leeuwen, 1996) and mediated through shared social and culturally situated activity (Vygotsky, 1935/1978). As theories of semiotics are advanced, prior work has focused on monomodal domains of the various art forms and articulated the associated grammars of each individual system of communication (i.e., visual, auditory, etc.). More recently, researchers are promoting a multimodal view of literacy that is key to understanding newer art forms (Kress & van Leeuwen, 2001). These efforts broaden our conceptions to include a theory of "multimodal literacy" and what it might mean to "make meaning" across a range of modalities. Jewitt and Kress (2003) argue for two central practices in their theory of multimodal literacy, including "design thinking" as encapsulating the intentions of a designer in absentia of the materials and the "production thinking" emanating from those ideas in the materials. In sum, artists make sense of individual modalities with the ultimate goal of making connections between several different types of modalities. Arts engagement fosters the ability to translate one type of literacy to another.

Further, current research on language and literacy point to the ways in which various modes of communication have value in the larger social and cultural context. For example, each of the major art forms (e.g., dance, drama, music, and visual arts) can be seen as its own symbolic system of language, one capable of expressing a range of emotionality and communicating a rich set of ideas and understandings that is oftentimes unattainable through speech alone. Not surprisingly, this realization allows us to view the meaning making of disenfranchised groups in new ways (Baum, Owen, & Oreck, 1997; Catterall & Waldorf, 2000), which is especially true for young children and those with disabilities that were previously seen as illiterate or pre-literate. Young children's drawings, for example, can now be seen as efforts at meaning making and expression and can be used for thinking and reflection (Kress & van Leeuwen, 1996). Through seeing these acts as art in their own right (Gardner, 1980), we begin to recognize these acts as literate activities and can begin to understand how young children as well as all learners begin to read and write the world through artistic acts.

Moreover, artists shape and convey intellectual and emotional content in their artwork as well as to evoke intellectual and emotional responses in the viewer (Greene, 2001). As such, we begin to see that learning in the arts transcends the benefits for the artists and includes the audience in the learning process as well. This is because the art object presents opportunities for the audience to engage in learning. The arts naturally afford inclusive learning opportunities because arts tradition is deeply rooted in performance, ,which is a natural culmination of dance, drama, and music. Also, displaying the final product in the visual arts is a common culmination, which positions the audience as a primary motivating force in the arts (Sefton-Green, 1998).

As theories of constructionism would explain (Papert, 1980; Kafai, 2006), the process of producing a work of art engages the artist in an iterative exploration of ideas and emotions as the work proceeds in a meta-cognitive manner. During this process, the artist learns to refine aesthetic sensibilities and build knowledge about materials, while connecting to other disciplines. For example, figurative drawing engages an artist in further understanding human anatomy. The fact that the arts give use new ways to read and write the world, has spurred a flurry of research aimed at the arts and the transfer of learning to other traditional academic areas such as mathematics, spatial reasoning, and oral and written language acquisition (Winner & Hetland, 2000; Deasy, 2002). Studies and commentaries in the publication, *Critical Links* (Deasy, 2002), have accumulated support that the arts and oral and written language share interrelated physical and symbolic processes, an area of research that could be further explored in the learning and cognitive sciences.

Relationships between the arts and literacy and language development are found across all of the visual and performing art forms, though the research is currently most robust in music and drama. For example, music features a symbol system that shares fundamental characteristics with language. At its core, music can be seen as decoding and encoding procedures that have syntactic and expressive structures (Scripp, 2002). In a similar manner, Catterall found that dramatic enactments enhance youths' abilities to comprehend texts, identify characters, and understand character motivations (2002). Studies also indicate that dramatic activities promote both writing proficiency and prolixity in generating written material. Further research is needed into other artistic forms, particular dance and the visual arts, as very little current research exists in this area. Additionally, learning scientists could help to unpack the mechanisms at work in the connections between the arts and literacy. For example, math educators call for more research on learning and aesthetics in order to develop a mathematics aesthetic, which evokes awareness of the beauty in mathematical ideas (e.g., Sinclair & Crespo, 2006). Across these studies, we see that the arts set the stage for learning a unique language and communicating with others in the world at large. Further inquiry into theorizing such literacies is needed, with the learning sciences uniquely poised to contribute knowledge on design-based research, assessment, and situated learning.

## Discussion & Conclusion

We have pointed to ways the field of arts education could benefit from research in the learning sciences as well as the ways in which the arts alert us to an understanding that is under-represented in the learning sciences. There are a range of cognitive, social, and cultural capacities engaged by learning in the arts. The following summarizes five opportunities for future research. First, research is needed to examine how youth develop such knowledge-in-action as well as the disposition to see the world through the lens of aesthetics, as we unravel how youth wrestle with ideas, materials, and meanings in the arts (Fiske, 2000). Second, there is a need to better understand how the various art forms uniquely impact the learning experience. For example, in what ways is learning in music distinct from other ways of representing ideas, such as dance or visual arts? Deasy adds that "finding alignment among the ways in which the study of different artistic forms demands and nurtures complex thinking has significance for the development of comprehensive arts programs and for our understanding of the nature of thought in arts learning" (2002, p. 6). Third, further research is needed on the design of learning environments in the arts. Better understanding of the pedagogical approaches and classroom contexts that support learning in the arts is needed (Horowitz & Webb-Dempsey, 2002). Fourth, there is also a need to define and measure "arts learning" (Catterall, 2002). Current studies do not unpack the extent and quality of the learning experience in the arts. Consequently, we know little about the specific qualities of arts learning that contribute to the gains aforementioned in this paper. As a result, it is difficult to qualitatively compare arts programs and there is an assumption that all arts education is of similar quality. This does not contribute to our understanding of how variations in learning in the arts account for variations in learning outcomes. Lastly, further research is needed in traditional areas of the learning sciences to explore the role of arts and aesthetics in our understanding of design, representations, and research methodologies, among a host of other domains in the learning sciences.

## References

Alexander, K., & Day, M. (Eds.). (1992). Discipline-based art education: A curriculum sampler. Los Angeles: Getty Trust Publications.

Allen, P. B. (1995). *Art is a way of knowing*. New York: Shambhala Publications, Inc.

Arts Education Partnership. (2004). The arts and education: New opportunities for research. Washington, DC: Author.

Barone, T. (2008). How arts-based research can change minds. In M. Cahnmann-Taylor & R.

Baum, S., Owen, S., & Oreck, B. (1997). Transferring individual self-regulation process from arts to academics. *Arts Education Policy Reveiw*, 98(4), 32-39.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences*, *2*(2), 141-178.

Cahnmann-Taylor, M., and Siegesmund, R. (Eds.). (2008). *Arts-based research in education: Foundations for practice*. New York, NY: Routledge.

Catterall, J. S. (2002). *Critical links: Learning in the arts and social development*. Washington, DC: The Arts Education Partnership.

Catterall, J. S. (2009). *Doing well and doing good by doing art*. Los Angeles: I-Group Books.

Catterall, J. S. (2004). The arts and education: New opportunities for research. Washington, DC: The Arts Education Partnership.

Catterall, J. S. and Peppler, K. A. (2007). Learning in the visual arts and the worldviews of young children. *Cambridge Journal of Education*, 37(4):543–560.

Catterall, J. S., & Waldorf, L. (2000). Chicago arts partnerships in education summary evaluation. In E. B. Fiske (Ed.), *Champions of change: The impact of the arts on learning* (47-62). Washington, DC: The President's Committee on the Arts and the Humanities, Arts Education Partnership.

Christians, C. G. (2005). Ethics and politics in qualitative research. In N.K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed.).(139-164). Thousand Oaks, CA: Sage Publications, Inc.

Cole, A. L., & Knowles, J. G. (Eds.). (2001*). Lives in context: The art of life history research.* Walnut Creek, CA: Alta Mira Press.

Danish, J. A. (2009). *BeeSign: A design experiment to teach Kindergarten and first grade students about honeybees from a complex systems perspective.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Deasy, R. (Ed.). (2002). *Critical links: Learning in the arts and student academic and social development.* Washington, DC: Arts Education Partnership.

Desantis, K. and Housen, A. (2000). A brief guide to developmental theory and aesthetic development. *Visual Understanding in Education*. New York, NY.

Dewey, J. (1980). *Art as experience*. New York: The Penguin Putnam Inc. (Original work published 1934)

Efland, A. D. (1990). *A history of art education: Intellectual and social currents in teaching the visual arts*. New York, NY: Teachers College Press.

Eisner, E. W. (2002). *The Arts and the Creation of Mind*. New Haven: Yale University Press.

Eisner, E. (2008). *Persistent tensions in arts-based research*, chapter 2, pp. 16–27. New York, NY: Routledge.

Fiske, E. B. (Ed.). (2000). *Champions of change: The impact of the arts on learning*. Washington, DC: The President's Committee on the Arts and the Humanities, Arts Education Partnership.

Franklin, B. (1931). Proposed hints for an academy. In J. Woody (Ed.), The educational views of Benjamin Franklin. New York: McGraw Hill. (Original work published 1749)

Friedman, M. (1989). Feminism and modern friendship. *Ethics*, 99(2), 275-290.

Freedman, K. (2003). *Teaching visual culture: Curriculum, aesthetics,and the social life of art*. New York, NY: Teachers College Press.

Gardner, H. (1980). *Artful scribbles*. New York: Basic Books.

Gardner, H. (1991). *The unschooled mind*. New York: Basic Books.

Greene, M. (1995). *Art and imagination*. San Francisco, CA: John Wiley & Sons, Inc.

Greene, M. (1995). *Releasing the Imagination: Essays on education, the arts, and social change.* San Francisco: Jossey-Bass.

Greene, M. (2001). *Variations on a blue guitar: The Lincoln Center lectures on aesthetic education*. New York: Teachers College.
Heath, S. B., & Soep, E. (1998). Living the arts through language+ learning. *Americans for the Arts Monographs* (Novemenber), 2.7, 1-18.
Hetland, L., Winner, E., Veenema, S., and Sheridan, K. M. (2007). *Studio thinking: The real benefits of visual arts education*. New York, NY: Teachers College, Columbia University.
Hill, A. (1951). *Painting out illness*. London: Williams and Norgate
Horowitz, R. & Webb-Dempsey, J. (2002). Promising signs of positive effects: Lessons from the multi-arts studies, in R. Deasy (Ed.), Critical links: Learning in the arts and student academic and social development. Washington, DC: Arts Education Partnership.
Irwin, R. L. (2004). A/r/tography: A metonymic metissage. In R.L. Irwin & A. de Cosson (Eds.), *a/r/tography: Rendering self through arts-based living inquiry*. Vancouver, Canada: Pacific Educational Press.
Irwin, R. L. and de Cosson, A. (2004). *a/r/tography: Rendering through arts-based living inquiry*. Vancouver, Canada: Pacific Educational Press.
Jewitt, C. & Kress, G. (2003). *Multimodal literacy*. New York: Peter Lang.
Kafai, Y.B. (2006). Constructionism. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences.* Cambridge, MA: Cambridge University Press.
Kaplan, S. Bardwell, L. V., & Slakter, D. A. (1993). The museum as a restorative experience. *Environment and Behavior*, 25, 725-742.
Katz, L. G. and Chard, S. C. (1997). Documentation: the Reggio Emilia approach. *Principal*, 76:16–17.
Kennedy, J.R. (2002). The effects of musical performance, rational emotive therapy and vicarious experiences on the self-efficacy and self-esteem or juvenile delinquents and disadvantaged children. In R. Deasy (Ed.), *Critical links: Learning in the arts*. Washington, DC: Arts Education Partnership.
Kress, G. , & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London: Reutledge.
Kress, G. & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.
Lave, J. (1987). *Cognition in practice*. New York: Cambridge University Press.
Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
Leavy, P. (2009). *Method meets art: Arts-based research practice*. New York, NY: The Guilford Press.
Lynch, M. (1988). The Externalized Retina: Selection and Mathematization in the Visual Documentation of Objects in the Life Sciences. *Human Studies, 11*, 201-234.
Malchioldi, (1999). *Medical art therapy with children*. London: Jessica Kingsley Publishers, Ltd.
Maslow, A.H. (1943). A theory of human motivation. *Psychology Review*, 50(4), 370-396.
Mcniff, S. (2004). *Art heals: How creativity cures the soul*. Boston, MA: Shambhala Publications, Inc.
Noblit, G. W., & Corbet, D. (2001). North Carolina charter school evaluation report. Evaluation Section, Division of Accountablity Services, North Carolina Department of Public Instruction. Submitted to State Board of Education.
Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.
Penny, S. (2000). Agents as artworks and agent design as artistic practice. In K. Dautenhahn (Ed.), *Human cognition and social agent technology* (pp.395-414). Philadelphia: John Benjaminis Publishing.
Pink, S. (2001). *Doing visual ethnography: Images, media, and representation in research*. London: Sage.
Rohd, M. (1998).Theatre for community, conflict & dialogue: The hope is vital training manual. Portsmouth, NH: Heinemann.
Saldana, J. (2008*). The drama and poetry of qualitative method. In M. Cahnmann-Taylor & R. Siegesmund (Eds.). Arts-based research in education: Foundations for Practice* (pp. 220-227). New York, NY: Routledge.
Scripp, L. (2002). An overview of research on music and learning. In R. Deasy (Ed.), *Critical links: Learning in the arts and student academic and social development*. Washington, DC: Arts Education Partnership, 132-136
Sefton-Green, J. (Ed.). (1998). *Digital diversions: Youth culture in the age of multimedia*. London: UCL Press.
Sinclair, N., & Crespo, S. (2006). What makes a good problem: An aesthetic lens. *International Group for Psychology of Mathematics Education*, 5, 129-136.
Springgay, S. (2004). *Inside the visible: Youth understandings of body knowledge through touch*. Unpublished doctoral dissertation, University of British Columbia, Vancouver.
Sullivan, G. (2005). *Art practice as research: Inquiry in the visual arts*. Thousand Oaks, CA: SAGE Publications.
Vygotsky, L. (1935/1978). *Mind in society*. Cambridge, MA: Harvard University Press.
Winner, E., & Hetland, L. (Eds.). (2000). Special Issue: The Arts and Academic Achievement: What the evidence shows. *Journal of Aesthetic Education*, 34(3-4), 1-307.

# SPACE AND TIME IN CLASSROOM NETWORKS: MAPPING CONCEPTUAL DOMAINS IN MATHEMATICS THROUGH COLLECTIVE ACTIVITY STRUCTURES

Tobin White, UC Davis
Corey Brady, University of Virginia

**Abstract:** This paper reports on a design-based research project in which we are developing collaborative activities for classroom networks. Research in generative design has explored classroom activities that analogize the many participants in a classroom with an often-infinite family of mathematical objects. By contrast, our project focuses on classroom activity structures that support interactions among two to four students, each aligned with different components of a shared mathematical object. In each case, these network-based relationships among students are intended to serve as resources to support learners' efforts to jointly navigate the conceptual territory delineated by their corresponding mathematical relationships in the space of the network. We compare the kinds of exploration of mathematical terrain respectively supported by whole-class and small group activities in classroom networks in order to examine the distinctive learning opportunities provided by each, and consider the potential of these different activity structures for providing mutually supportive instructional experiences.

## Introduction

Simultaneously accounting for the cognitive and situative dimensions of learning remains an important and persistent challenge for contemporary mathematics education (Greeno, 1997; Sfard, 1998). Prior work along these lines highlights the importance of attending to intersections between discipline-specific content and practices in mathematics on the one hand, and social norms and interactions across different classroom activity structures on the other. For example, in a seminal account of the links between classroom interactions and mathematical activity, Yackel and Cobb (1996) observed that social structures, in the form of the collective, local norms for participating in mathematical discourse, serve as resources for structuring classroom mathematical practices. Following a similar insight in the reverse direction, research on generative design (e.g., Stroup, Ares and Hurford, 2005) has pointed to ways that mathematical structures can serve as resources for structuring social activity through a classroom network of handheld calculators. The intersection between individual students' engagement with mathematical objects in the private space of their respective devices and the collectively constructed artifacts visible in classroom networks provides rich possibilities for exploring this dynamic between conceptual and social.

In a similar vein, this paper reports on a design-based project in which we develop collaborative activities for classroom networks in order to investigate intersections between conceptual and social dimensions of mathematics learning. As in generative design, our own approach uses the mapping between student participants in a classroom group and mathematical objects in a shared virtual space to make important mathematical relationships and properties of these objects salient. Generative designs analogize the many participants in a whole-class activity with an often-infinite family of mathematical objects. For example, each student might be associated with a different function in a family of functions; a different point in a locus of points; or a different expression among a class of equivalent expressions (Stroup, Ares, Hurford & Lesh, 2007). By contrast, our project focuses on network designs and classroom activity structures that support collaborative interactions and investigations among groups of two to four students. These designs align each student participant in a small group with one of a small number of components of a single shared mathematical object, such as different representations of a common function (White, 2006).

In each case, these network-based relationships among students are intended to serve as resources to support learners' efforts to jointly navigate the conceptual territory delineated by their corresponding mathematical relationships in the space of the network. We see the forms of virtual exploration of interrelated graphical and symbolic terrain afforded by these designs as offering a visual example of Greeno's (1991) proposal that knowledge of a domain is analogous to familiarity with an environment. Our aim in this paper is to consider the ways these different activity structures for classroom networks (at whole-class and small group levels) might collectively map a corresponding conceptual territory in mathematics. In other words, we compare the kinds of exploration of mathematical terrain respectively supported by whole-class and small group activities in classroom networks in order to examine the distinctive learning opportunities provided by each, and to consider the potential of these different activity structures for providing mutually supportive instructional experiences.

## Collective Activity in Classroom Networks

A rich body of prior research has explored ways of using of classroom networks to support whole-class interactions (Hegedus & Kaput, 2004; Stroup, Ares & Hurford, 2005). A major strand of this work emphasizes using the sociocultural diversity of the class group as an engine for exploring the possible variation among families and other collections of mathematical objects (Stroup, Ares, Hurford & Lesh, 2007). For instance, in studying the notion of algebraic equivalence, students might be asked to invent and contribute expressions equivalent to a given expression (e.g., 4x). The resulting collection of expressions that are the same as 4x indicates the breadth of the space of this *equivalence class* of expressions (Stroup, Ares & Hurford, 2005). Students make use of key algebraic skills as they work to populate this space, "unsimplifying" the given expression. Because of the diversity of ideas present in virtually any classroom-sized group of participants, the range of student contributions will reliably instantiate key concepts such as the distributive law and identity and inverse properties of multiplication and addition (see, e.g., Davis, 2007).

Thus, this strand of research taps into the protean sea of the classroom group, confident that it contains sufficient variability and diversity to represent the mathematical universe adequately for pedagogical purposes. The size of the classroom group appears as a critical asset in these classrooms, as it serves as the effective guarantee that the hypothesis of sufficient-diversity will hold. We interpret activities that rely on the diversity of student contributions in this way to be using the size and diversity of the classroom group to concretize the concept of one of several kinds of *infinity*. Indeed, the representative activities of this genre tend to focus on aggregate mathematical objects such as families of functions, loci of points, or instances of geometrical relationships—all consisting of an infinite set of constituent objects. In each case, a critical learning objective is the appreciation of the nature of this infinity and of structures within it. The design of the activities allows the student to grasp the infinite family, locus, or range of figures without losing contact with the specificity of the examples—her own and those of her peers—that serve to embody the infinite aggregate.

While these instances of infinity represent a rich and varied set of mathematical topics, they do not span the full array of ideas central to the discipline. Indeed, many core mathematical objects under study in the k-12 curriculum are composed of sets of just a few coordinated sub-objects: lines are uniquely determined by two points, quadratic expressions are comprised of three distinct monomial terms, equations compare two distinct algebraic expressions, functions are commonly represented through one of three modes—symbols, graphs or tables. Additionally, many of the conceptual challenges that face students as they grapple with these concepts consist of difficulties with the coordinated significance of these sub-objects: how to determine the slope of a line, to combine polynomial terms and simplify expressions, to solve equations, to interpret relations among representations.

With these concepts and challenges in mind, our own approach uses network links among student devices in order to align pairs or small groups of students with correspondingly small sets of mathematical objects. As in the case of analogizing the many students in a whole class to infinitely many related mathematical objects, these small group designs likewise seek to capitalize on network-defined social relationships to make corresponding mathematical relationships salient in learners' activity. To that end, our designs focus on collaborative learning tasks that are centered on collective mathematical objects that participants in a pair or small group must jointly manipulate through their networked devices. Such objects are collective to the extent that they or their attributes appear—and change—simultaneously on the devices of multiple students, and that they appear in a shared display as a consequence of contributions from multiple students. These designs map the students in a small group to sub-objects in a coordinated set: each learner manipulates a linked point to collectively form a curve in a Cartesian plane, or transforms a respective side of an algebraic equation, or combines different-ordered terms or enacts different binomial operations among polynomial expressions, or examines a different representation of the same function. Broadly, we aim to structure tasks around these collective mathematical objects in order to make the successful solving of problems dependent on contributions from and coordination between all participants in a small group. Below, we describe one such design in detail, and present an analysis of the kinds of student mathematical exploration supported by this environment.

## The *Graphing in Groups* Design

The collaborative mathematics activities described in this paper were created using the NetLogo modeling environment (Wilensky, 1999) and HubNet architecture (Wilensky & Stroup, 1999) in concert with classroom sets of Texas Instruments graphing calculators connected through a TI-Navigator™ network. In the activity design called

*Graphing in Groups*, teams of two or more students share collective responsibility for jointly manipulating the graph of a function (for two students, a linear function; for three students, a quadratic). Each calculator displays a graphing window and allows the student to adjust the coordinates of a point graphed within that window using directional arrow keys on the calculator (Figure 1). The coordinate points inputted through the calculators of each member in a student pair, along with the line uniquely determined by those respective points, are displayed together in a single graphing window on the classroom server, creating a shared mathematical space for the members of a group (Figure 2). A grid composed of several groups' graphing windows and projected at the front of the classroom provides a collective display of all these small group-level graphs (Figure 3).

In the activities for the present study, each student's point was paired with that of a single partner so that their respective coordinates collectively defined a line. As students moved the individual point controlled by their respective graphing calculators, they could choose to "mark" the current coordinate location of that point by pressing a calculator key. Once both students in a pair marked points, the corresponding line appeared in their group's graphing window, along with an equation for that line in slope-intercept form (i.e., y=mx+b). Students could continue to move their respective points freely, and any time either student in a pair pressed the "mark" key again, the graph and equation automatically changed accordingly in the public display. In other words, the students controlled both an individual object, in the form of their respective points, and a collective object in the form of a line.

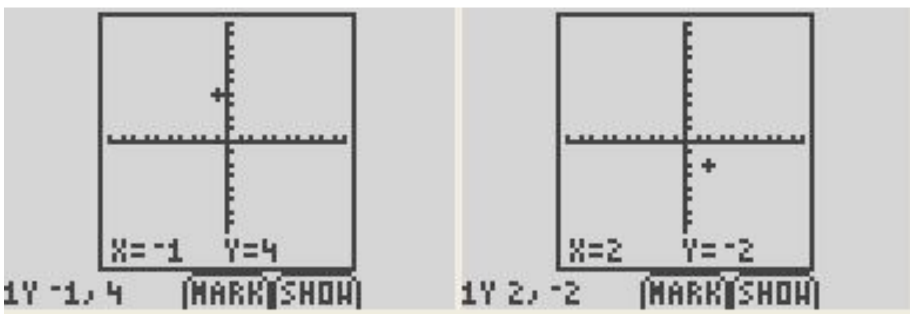



Figure 1. Two student calculator screens.

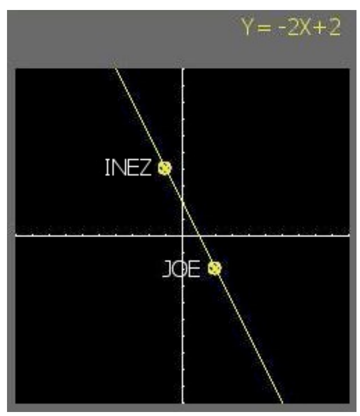



Figure 2. Two student points in a group-level graph.

## Method

The *Graphing in Groups* design served as the basis for a three-week unit on linear graphing implemented in three ninth grade Algebra I classes, each taught by a different teacher. Over a series of several class sessions, students participated in a variety of *Graphing in Groups* activities intended to complement more conventional lessons on linear equations and their graphs. The topics addressed during this unit included slope, x- and y-intercepts, the point-slope and slope-intercept forms of linear equations, and parallel and perpendicular lines. Some of the *Graphing in Groups* activities conducted during the unit were teacher-led explorations and whole-class discussions of the relations among lines generated by different pairs. Other activities were problem-solving challenges assigned to student pairs. These problem-solving tasks generally involved asking student pairs to move their respective points so as to create a line with a certain characteristic or set of characteristics: a slope of three, a y-intercept of four, a negative slope, a line parallel to a teacher-generated line, etc. Such tasks were typically closed in the sense that each pair worked toward a single solution, but open with regard to the coordinates each student might assume and the process by which they identified that solution. In each case, the instructional focus was on framing these challenges in terms of pairs' collective lines, so that completing that task required students to consider that line and its equation in relation to the coordinates of their respective points.

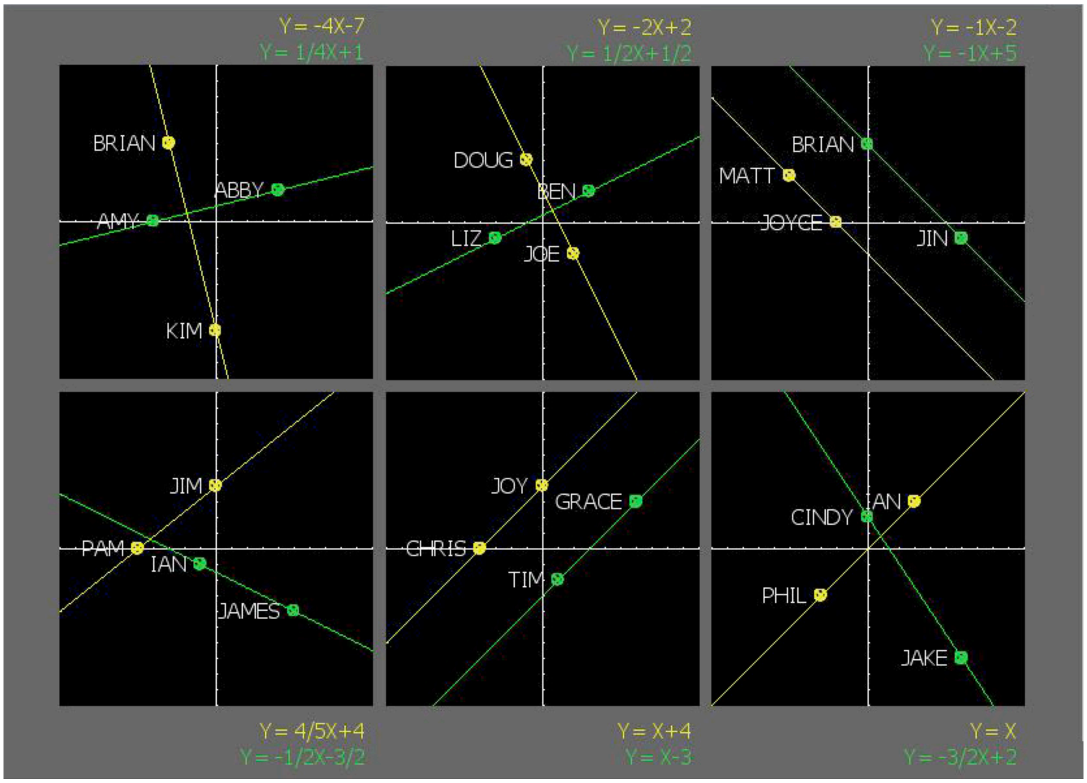



Figure 3. *Graphing in Groups* whole-class display.

Students were assigned to pairs by the teacher, and remained in the same pair throughout the unit. Two to three student pairs in each class were selected, on the basis of informed consent, as focus groups and videotaped during all collaborative activities throughout the unit. An additional camera captured the projected server display at the front of the room, as well as whole-class discussions and other instructional interactions. Server logs captured coordinates entered by the students as they manipulated their respective points to complete the collaborative tasks. Analysis of these data focused on the ways students individually and collectively interpreted the mathematical objects (i.e. points, lines and equations) and on their relations in this environment as they developed approaches to solving problem-solving tasks. Video segments of six student pairs working on between 13 and 19 tasks each were transcribed and examined alongside time-synched video of each group's shared graphing window from the server display. Students typically worked three to four minutes before completing or stopping each task; the longest sessions ran approximately 10 minutes and the shortest lasted about 30 seconds. Each transcript was then annotated with detailed descriptions of the ways students moved and marked their points. The first author and another researcher developed two different coding schemes for analyzing these annotated transcripts for each task episode: one focused on problem-solving strategies and the other on interactions between students.

Below, we analyze an episode from the video data in order to illustrate the ways learners engaged with mathematical objects and relationships in the context of this network-supported activity for student pairs. In particular, we examine the ways students explored and interpreted the relations between their respective points and a collective line, and between the graph and the equation of that line, over the course of a problem-solving task. This episode is typical of the broader data set relative to the analytic coding scheme; both the problem-solving strategies employed and kinds of student interactions occurred frequently in other tasks and among other pairs. We have selected it for presentation in the current paper because it provides a clear example of students' explorations of this mathematical domain as they undertook these problem-solving tasks, and because it features a task reminiscent of the "make expressions that are the same as 4x" example of generative whole-group activity in classroom networks.

## Analysis

This episode finds two students, Priyana and Kate, in the midst of efforts to locate their respective points at coordinates that would collectively form a line with equation y=4x-1. Just prior to the excerpt below, the students had worked for a few minutes by moving each of their respective points before realizing that they could establish a line with the correct y-intercept by having Kate position her point at -1 on the y-axis. In the moments that followed, Kate kept her point stationary at the y-intercept, watching and offering commentary and suggestions as Priyana moved her own point in a series of attempts to construct a line with slope 4. To do so, she began taking a series of paired horizontal and vertical unitary steps, marking and examining the results of each move:

1. [Kate has just marked at (0,-1) to make y=(5/3)x-1. From (3,4), Priyana moves up and right to (4,5), marks to form y=(3/2)x-1.]
2. Kate: That slope has to be the same, so…what was the slope before, a 4?
3. Priyana: [Moves up and right again to (5,6)] A negative 4. [Presses mark to form y=(7/5)x-1]. I mean a 4. Ok…[moves up and right again to (6,7)]
4. Kate: Ok, so…
5. Priyana: [Presses mark at (6,7) to form y=(4/3)x-1. Moves right and up to (7,8) and presses mark to form y=(9/7)x-1. Moves up and right again to (8,9), where the label of her point partially covers the pair's equation]
6. Kate: [Trying to read the obscured equation] What does it say? [Priyana now marks at (8,9) to form y=(5/4)x-1] Can you go down more?
7. Priyana: Yeah, maybe I should go down. [Moves down five units and left five to (3,4), then back one unit right to (4,4), presses mark. Equation does not change. Looks down at her calculator and then up at the screen again, moves down and left to (3,3) and marks to form y=(4/3)x-1. Moves down and left again to (2,2) and marks to form y=(3/2)x-1. Moves down and left again to (1,1) and marks to form y=2x-1.] Oh, there it is.

This segment illustrates the kind of incremental movements in a graphical space that students often used in order to investigate variations in a corresponding algebraic expression in this environment. Priyana began with five successive moves up one unit and over one (Figure 4a), marking and observing the resulting slope each time (lines 1-6). She used these paired unitary vertical and horizontal steps—one up and one over—to make incremental adjustments to the slope of the line, watching the resulting changes in the collective display and reacting to each with soft vocalizations (e.g. "Ok" in line 3) or pursed lips before adjusting again. In each case, she used minor adjustments to the graph as a means of gradually adjusting the value of the linear coefficient in the equation. Importantly, though, she did not simply make the smallest possible adjustment in *her position*—one unit up *or* one over. Rather, taking an incremental step in each dimension before marking to form a new line allowed her to make the smallest possible adjustment to *their line*.

The next short series of movements further highlights the students' careful efforts to alternately maintain or only minimally rotate the current line. As Priyana's point moved toward the upper right corner of the graphing window, the label on her point actually obscured part of the pair's equation (line 5). Kate, who was closely watching each move, asked Priyana to "go down more" so that she could read the equation clearly (line 6). Priyana, who was quickly running out of screen room, agreed that "maybe (she) should go down" (line 7). She retraced her steps in reverse, moving down five units and left five and then briefly pausing as though to consider returning to (3,4), the point at which she had started the sequence of moves that opened this episode. Instead, she moved one unit back to the right to position herself on the same line before marking at (4,4), and so leaving the current line unchanged. Again, these moves appeared carefully selected to make only minimal changes to the current line—to seek out the correct parameter through incremental adjustments. Priyana thus abandoned her initial sequence of similar moves only when display limitations interfered, and worked carefully to minimize changes in the line even as she adopted a new line of exploration. Continuing to display the same preference for relatively small variations in slope, Priyana began a new series of paired unitary horizontal and vertical steps, this time down and to the left from (4,4) (Figure 4b). Three such steps established a slope of two, at which point she noted the appearance of an integer slope value (line 7). At this point, Kate and Priyana briefly disagreed about how to proceed:

8. Kate: Go up like… go up a little bit. [Ignoring Kate's instruction, Priyana moves down to x-axis, left to origin, and up to (0,2), all without marking] It's supposed to be 4.
9. Priyana: [Continues circling her point around her last mark—two units right to (2,2), then back to x-axis, then left toward origin] Yeah, maybe if I keep going down, like this… [briefly brings her point to rest on the origin, as if to mark there, one unit left and one down from her last mark, but then moves down again to sit on top of

Kate's point at (0,-1) and presses mark so that the pair's line disappears]. That's not right. [moves back up to (1,1) and marks to again form y=2x-1].

10. Kate: [As Priyana moves up and over one to (2,2), but does not mark, and instead moves back to (1,2)] Go up, like, little by little.
11. Priyana: [marks at (1,2), forming y=3x-1]. There you go. Oh, hold on, I'll just do one more. [Moves up to (1,3)]
12. Kate: [reading the y-intercept from the equation] That's a negative 1 though.
13. Priyana: [Marks at (1,3) to form y=4x-1] There you go.

This excerpt highlights an important contrast between Priyana's continuing incremental exploration and Kate's recognition of a more direct path toward the desired line. Reacting to the newly established slope of two, Kate prompted Priyana to "go up a little bit" (line 8). Ignoring the instruction, Priyana wavered, appearing to consider movement in all possible directions as she traced a square path around her last mark at (1,1) (Figure 4c). Wondering aloud, "maybe if I keep going down, like this," she briefly came to a stop at the origin and directly above Kate's point, where a mark would have continued her previous pattern of paired unitary steps down and left, but also formed a line with undefined slope (line 9). While Kate now recognized that moving Priyana's point up would bring about the desired increase in the value of the slope, Priyana had to this point continued relying on incremental steps in the graphical space in order to feel her way about in search of a solution. However, the fact that she chose not to mark here at the origin indicates that she recognized she was reaching the limits of this approach. Indeed, as she instead moved down to mark on top of Kate's point and briefly erase the line, she immediately realized that her new location was "not right" and returned to her previous mark at (1,1) (line 9). Kate again urged her to "go up…little by little" (line 10). Finally acquiescing, Priyana marked one unit up at (1,2). When the slope changed from two to three, Priyana clearly recognized that they were now moving toward a line with the correct slope, and that she needed to "just do one more" similar step, quickly marking at (3,1) to form the target slope and line (Figure 4d).

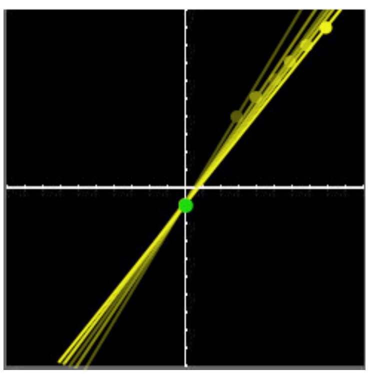


Figure 4a.

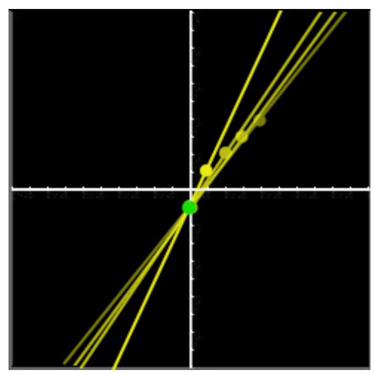


Figure 4b.

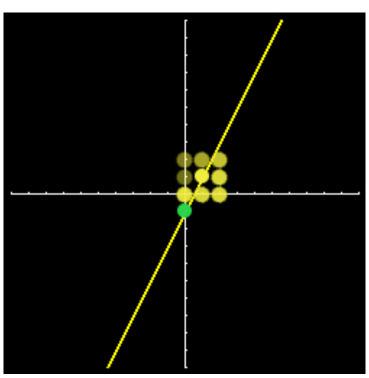


Figure 4c.

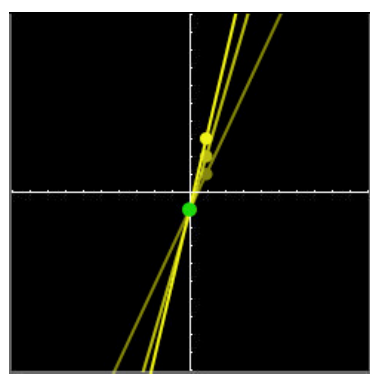


Figure 4d.

Throughout this sequence, Priyana moved from incrementally adjusting the slope downward and away from the correct value using paired unitary horizontal and vertical steps, to adjusting the slope in the same manner but upward and toward the correct value, to increasing the slope from two to the target value of four by making unitary vertical steps upward. In the first two instances, neither she nor Kate appeared to recognize the relationship between the steepness of the line and the value of the slope. In the absence of a clear target toward which to adjust the line, Priyana focused on making small and systematic changes away from the current state, first in one direction and then in another, until she established a parameter value she and Kate were better able to interpret. That moment arrived when the slope went from a series of fractions to an integer; at that point Kate urged her to “go up a little bit” because “it’s supposed to be four” (line 8), clearly associating upward vertical change in Priyana’s point location with increases in the slope. By contrast, Priyana initially appeared to interpret the achievement of an integer close to the target as an indication of the fruitfulness of her unitary moves down and left. When applying this interpretation proved unsuccessful, she instead followed Kate’s advice, and appeared to quickly come to the same insight about the relations between her vertical moves and the slope of the line.

Priyana’s transition from movements that indicate some confusion about the relationship between the graphical and the symbolic to those that reflect a clearer and more correct interpretation of that relationship suggest ways that successive, iterative movements in this space serve as resources for making meaning about slope. Both students appeared to be coming to clearer understanding of the relations between graphical movements and linear equations through real-time incremental adjustments over the course of the episode. Indeed, this episode in particular and our analysis of students’ strategic engagement with these tasks generally suggests that patterned, iterative sequences of incremental movements like Priyana’s here are particularly important resources for students’ efforts to navigate this mathematical terrain. Though these movements are sometimes exploratory and ad hoc, much of the time they indicate students’ clear expectation that these particular changes will bring about particular results. For example, students often articulated specific directions (“go up, like, little by little,” line 10) to one another or narrated their own moves (“I’ll just do one more,” line 11), suggesting that they were beginning to anticipate certain outcomes from these moves.

## Discussion

In the episode above, moving points in a graphical display allowed students to explore the mathematical space of linear functions with rational-number slopes. By systematically and incrementally varying coordinates, Priyana and Kate were able to feel their way about in the unfamiliar territory of fractionally-sloped lines, and to make use of a more familiar landmark when they encountered an integer value for slope. These kinds of exploratory movement typical of students’ engagement with the *Graphing in Groups* interface illustrate the ways that developing strategies for moving points and manipulating functions in these classroom network activities might be understood as learning to navigate a conceptual environment. In this sense, we see students’ joint exploration of virtual mathematical terrain as a productive alternative interpretation of the physical metaphor for a conceptual domain proposed by Greeno (1991).

From the standpoint of this analogy to conceptual environments, whole- and small-group activity structures may form complementary dimensions of a complex mathematical domain. The episode above suggests an illustrative contrast between the whole-class activity to “provide different expressions that are equivalent to 4x” and the *Graphing in Groups* challenge to “make a line with a slope of 4” from the raw materials of the two students’ respective coordinate locations. In the whole-class activity, the equivalent expressions are manifested as objects that fill out a conceptual *space*; in the *Graphing in Groups* activity, the small group’s solution is experienced as a process *in time* through a sequence of individual explorations and collaborative interactions. In *Graphing in Groups*, student movements of their individual points generate fluctuations and modifications of the collective graph, as the group moves incrementally towards a goal. Each student moves freely, but those movements shape and constrain the emergent activity of the group as teams explore the impact of incremental changes in their constituent sub-objects upon the nature and attributes of the collective, composite object. That collective activity is characterized not so much by attention to the family of objects that fill out a function space but rather by the manipulation over time of a single function through small perturbations—the exploration of *infinitesimal modification* rather than the unfolding of an infinite parameterization.

In generative designs for whole-group classroom network activity structures, diversity and individual expression in the classroom group serve as surrogates for infinity. In contrast, in the approach for small groups presented above, distinctive student actions take the form of infinitesimal variations in the group’s collective object. Because students are identified with complex or compound mathematical objects, with each group member given the role of a component sub-object, these variations demonstrate the dependence of the collective object on the characteristics of its components. Moreover, while the whole-class activity highlights the infinite variety of expressions equivalent to 4x, the pairs task emphasizes infinitesimal variations among linear terms increasingly close to 4x. In these ways, we see designs for whole- and small-group activity as respectively and jointly mapping spatial and temporal dimensions in a conceptual environment central to the broader curricular landscape of secondary mathematics.

## Conclusion

The relationship between the conceptual dimensions spanned by these small-group and (generative) whole-group activity structures suggests corresponding ways that they might be combined to form a complementary instructional sequence. While the analytic focus of this paper and the early phases of this design-based research project have been on mathematics learning opportunities afforded by small-group interactions, the *Graphing in Groups* collective display (Figure 3) also explicitly supports an additional layer of artifacts and instructional possibilities oriented toward the larger classroom group. For example, in a different *Graphing in Groups* task from the one depicted here, each pair of students creates a line with the same slope. While the activity at the level of each pair might unfold along the lines of the episode presented in this paper, the collection of groups in the class would typically together produce several representative members of a family of parallel lines. When the small-group phase of the activity is complete, each of the groups has produced a distinct solution to the challenge, and the aggregate begins to fill out the mathematical space of possibility defined by that challenge. This sequence aims to use the classroom network to integrate two levels of classroom activity, one in which the small group focus is on understanding the interrelations of points and lines through strategies of infinitesimal variation, and another in which the classroom group as a whole can see the infinite variety of possible lines that meet the activity's challenge criterion. Our current classroom research focus is on jointly examining these small-and whole-group levels of mathematical activity and instructional interaction in this hybrid environment. In these ways, we see small-group designs of the kind outlined in this paper as expanding the map of the social and mathematical terrain in classroom networks charted by prior research, and broadening the pedagogical modes across which those networks might support mathematics teaching and learning. Collectively, these whole and small-group activity structures highlight the rich potential of the social space of networks for mapping conceptual spaces in mathematics.

## References


Ares, N. (2008). Cultural practices in networked classroom learning environments. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 301-326.

Ares, N. & Stroup, W. (2004). Drawing on diverse social, cultural, and academic resources in technology-mediated classrooms. *Proceedings of the 26th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, 2, 837–844.

Davis, S. (2007). Generative Activities in Singapore: A Beginning. *Paper presented at the International Conference on the Teaching of Mathematical Modeling and Applications*, Bloomington, IN.

Greeno, J. (1991). Number sense as situated knowing in a conceptual domain. *Journal for Research in Mathematics Education, 22*(3), 170-218.

Greeno, J. (1997). On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5-17.

Hegedus, S. & Kaput, J. (2004). An introduction to the profound potential of connected algebra activities: Issues of representation, engagement and pedagogy. *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education, 3*, 129–136.

Sfard, A. (1998). On Two Metaphors for Learning and the Dangers of Choosing Just One. *Educational Researcher, 27*(2), 4-13.

Stroup, W., Ares, N., & Hurford, A. (2005). A dialectic analysis of generativity: Issues of network-supported design in mathematics and science. *Mathematical Thinking and Learning, 7*(3), 181-206.

Stroup, W., Ares, N., Hurford, A. & Lesh, R. (2007). Diversity-by-Design: The What, Why, and How of Generativity in Next-Generation Classroom Networks. In R. Lesh, E. Hamilton, & J. Kaput (Eds.), *Foundations for the Future in Mathematics Education* (pp. 367-394). Routledge.

White, T. (2006). Code talk: Student discourse and participation with networked handhelds. *International Journal of Computer-Supported Collaborative Learning, 1*(3), 359-382.

Wilensky, U. 1999. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

Wilensky, U. & Stroup, W. (1999). HubNet. http://ccl.northwestern.edu/netlogo/hubnet.html. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

Yackel, E. & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458-477.


## Acknowledgements

Support for this work was provided by an NSF Early Career Development Award, DRL-0747536, granted to the first author.

# Sequential Effects of High and Low Guidance on Children's Early Science Learning

Bryan J. Matlen & David Klahr
Department of Psychology, Carnegie Mellon University
bmatlen@cmu.edu, klahr@cmu.edu

**Abstract:** We describe a microgenetic approach aimed at examining the effect of different sequences of high vs low levels of instructional guidance on students' learning about concepts and procedures associated with simple experimental design – often called the "Control of Variables Strategy (CVS). Third-grade children were randomly assigned to one of four conditions in which CVS was taught via one of the four possible orderings of high or low instructional guidance: (High + High, High + Low, Low + High, and Low + Low). High guidance consisted of a combination of direct instruction and inquiry questions, whereas low guidance was comprised only of inquiry questions. Contrary to commonly held beliefs that high levels of guidance, and in particular, direct instruction, lead only to shallow learning, results show that repeated instances of high instructional guidance were advantageous for both learning and transfer of CVS. Moreover, the High + High group continued to demonstrate strong conceptual understanding of CVS relative to other groups five months after training. Overall the study suggests that strong amounts of instructional guidance are capable of exhibiting powerful effects on children's early science learning.

## Introduction

Effective instructional design is crucial to promoting learning and transfer. While great strides have been made within the field of instructional science over the past three decades, intense debate still exists over some fundamental issues (Kirschner, Sweller, & Clark, 2006; Klahr, 2009; Kuhn, 2007). For example, what, if any, amount of instructional guidance will maximize student learning? If there is an optimal level of guidance, where should it be placed during the course of instruction? The present study is aimed at addressing these fundamental issues facing the learning and instructional sciences in the context of elementary science education.

Of particular importance to science education is whether high levels of instructional guidance in the form of detailed, explicit instruction can produce robust learning (i.e. learning that transfers across extensive periods of time and to disparate tasks). Instruction at the extreme end of the explicitness spectrum – commonly known as "Direct Instruction" – has classically been construed as providing only short-term learning gains, leading to knowledge that is fragile and incapable of transferring to remote or "authentic" settings (Germann, Aram, & Burke, 1996; McDaniel & Schlager, 1990; Koedinger & Aleven, 2007; Kuhn & Dean, 2005). Direct instruction has also been criticized on the grounds that it allocates a passive role to learners, rather than promoting active student learning, whereas low levels of instructional guidance are more effective at inducing learners to construct knowledge on their own, thereby leading to more meaningful understanding (Kuhn & Dean, 2005). Indeed, Dean and Kuhn (2007) conclude from their investigation that, *"... direct instruction appears to be neither a necessary nor sufficient condition for robust acquisition or for maintenance over time*." (p. 394).

However, these claims are inconsistent with the results of several of our prior investigations of the effectiveness of direct instruction, in which we have repeatedly found positive learning gains in situations where we provided students with a high degree of instructional guidance: direct instruction coupled with inquiry questions. This pedagogical approach has consistently produced meaningful and significant performance gains immediately after training (Chen & Klahr, 1999; Toth, Klahr, & Chen, 2000) that have been sustained for at least a period of three years (Strand-Cary & Klahr, 2009).

In this paper we address a more nuanced version of the question about the optimal level of instructional guidance. Instead of simply asking "which is best?" we also consider the possibility that there may be an optimal temporal sequence of different levels of guidance. As Koedinger and Aleven (2007) have demonstrated in the context of cognitive tutors, each end of the high-to-low instructional guidance spectrum has its unique strengths and weaknesses. Accordingly, rather than pit one type of instruction against another, several recent studies have reframed the question in terms of how different *sequences* of high and low guidance lead to different levels of student learning. However, this research has yielded some contradictory, results. For example, Schwartz and Martin (2004) have suggested that high amounts of guidance are most effective after students attempt to invent solutions in minimally guided settings, an instructional method known as *Invention as Preparation for Future Learning* (IPL). In their study, students who invented original formulas for calculating variance benefited more from subsequent direct instruction than students who initially received instruction and then practiced applying equations of variance. The notion that students may initially fail to produce correct solutions in minimally guided settings, but that this failure can serve as a ‘productive’ learning experience when followed by instructional support (i.e. termed *productive failures*), has also been reported by Kapur (2008; In press). These studies, however, seem in contrast to research that have observed an *Expertise Reversal* effect as learners acquire knowledge (Kalyuga, 2007). According to this latter perspective, since working memory is limited, novices need high levels of guidance in order to ameliorate the strong demands of cognitive processing required of them: As learners gain expertise, however, low guidance is thought to be most beneficial because it reduces the amount of redundant information presented. In support of this, Kalyuga and colleagues have provided empirical evidence that novices benefit more from viewing modeled solution steps and that, as they gain domain expertise, they learn more from engaging in unstructured practice problems (Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Kalyuga & Sweller, 2004). Overall, the implications these approaches suggest for sequencing instruction appear to be incongruous.

Given the stark disagreement in the literature surrounding not only instructional sequence, but also the utility of providing high guidance in the form of direct instruction, the present study has two inter-related goals. First, we test the claim that strongly guided instruction is insufficient for promoting robust understanding (Dean & Kuhn, 2007). To this end, we contrast the relative effectiveness of 1) High Guidance, in which students were provided inquiry questions and direct instruction, and 2) Low Guidance, in which we encouraged learning through the provision of inquiry questions, while purposefully withholding direct instruction. Our second goal is to examine the efficacy of varying sequences of these forms of instruction. Thus, while simple contrasts of high vs low guidance have been used in earlier work, in the present study, we contrast all possible orderings of these two instructional approaches across two separate training sessions, leading to four experimental conditions (i.e. High + High, High + Low, Low + High, and Low + Low). By contrasting these four conditions we are able to assess whether there indeed exist certain progressions of high and low guidance that are more effective in promoting student learning than multiple sessions of high or low guidance by themselves.

We address these issues in the context of teaching children about the Control of Variables Strategy (CVS) for scientific experimentation. CVS is the procedure used to create unconfounded experiments by keeping the values of all factors the same while changing only the variable of interest in order to determine whether or not that factor is causal with respect to the experimental outcome. The procedures and concepts associated with CVS have become increasingly influential in state standardized tests in elementary school and appear on almost all state standards as well as the National Science Standards for elementary school students (Klahr & Li, 2005; National Research Council, 1996). Despite CVS's importance, however, children throughout the elementary school years often demonstrate difficulty in applying it to relevant situations (Chen & Klahr, 1999; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995), making it an ideal domain – for both theoretical and practical reasons – in which to explore these research questions.

## Method

### Participants
Forty-two third grade children (22 girls, 20 boys, $M$ = 9.03 years, $SD$ = .33 years) from three middle-class Pittsburgh elementary schools participated in the study. Children were randomly assigned to one of the four experimental conditions.

### Design
The overall design was microgenetic in nature (Siegler & Crowley, 1991), and consisted of four experimental conditions and 10 test phases. The first eight test phases were conducted approximately one week apart from each other, whereas the final two test phases occurred five months later. All phases were identical across each condition, with the exception of phases 3 and 4, where the experimental treatment varied by condition across two training sessions: (i) High Guidance followed by High Guidance (High + High), (ii) High Guidance followed by Low Guidance (High + Low), (iii) Low Guidance followed by High Guidance (Low + High), and (iv) Low Guidance followed by Low Guidance (Low + Low). Each test phase is described in detail, below.

### Procedure
#### Phase 1: Story Pre-test
The first phase consisted of a paper and pencil pre-test with six questions aimed at assessing children’s understanding of CVS. Three questions assessed children’s ability to create an unconfounded experiment (design questions), and three asked them to evaluate whether an existing experiment was unconfounded (evaluate questions). Children completed the Story pre-tests at their desks in their regular classrooms and were given as much time as they needed to finish the test. Story responses were scored as 1 if children correctly identified or designed experiments consistent with CVS, and all other responses were assigned a 0. Scores ranged from 0 to 6.

## Phase 2: Ramps Pre-test

Children were introduced to two physical ball and ramp apparatuses and were told that they would be designing experiments to test whether certain variables made a difference in how far the ball rolled. Children were told the four variables that could affect the outcome: the surface type (either fim or sif), ball type (bab or lof), steepness (steep or not steep) and starting gate (high or low) (Surface and ball type were given nonsense names to minimize children's expectations about their effect on the outcomes of their experiments). After children could identify these variables without the help of the experimenter, one of the variables was chosen at random and the child was asked to design an experiment to test whether that variable made a difference in the outcome. Children were allowed to set up an experiment and observed its outcome. No feedback was given. The procedure was repeated until all 4 of the variables were tested in a random order, and children were assigned a score of 1 if their set up varied the target variable and kept all other variables the same (CVS), and a score of 0 if the set up was of any other combination. Scores ranged from 0 to 4.

## Phase 3: Training 1

The same materials as the Ramps Pre-test (Phase 2) were used. At the start of Training 1, the child was shown the ball and ramp apparatus and asked to identify the four variables that might make a difference in how far the ball rolled. Once children had correctly identified the four variables, the child's condition determined what happened next.

(a) Children who experienced Low Guidance in Training 1 (i.e., those in the Low + High and Low + Low conditions) engaged in minimally guided discovery learning while being provided inquiry questions in order to scaffold their learning. Children were told that they would be setting up experiments to see what made a difference in how far the ball rolled down the ramps and, subsequently, they were asked to set up an experiment to test one of the four variables, chosen at random. Children were also asked two scaffolding questions to prompt their self-explanations and (minimally) guide them in the discovery process. The first question occurred after children set up their experiments, but before they viewed the experimental outcome. The experimenter asked the child why he/she had set up the experiment that way. The experimenter listened to the child's explanations and provided neutral feedback such as, "*okay*," or "*alright*" and then asked the child to roll the balls down the ramp. Upon observing the result, the child was asked whether they could tell for sure from the experiment whether the target variable made a difference in the outcome. This procedure was repeated for all four variables in a random order, and the child designed two experiments to test each variable, resulting in eight child-designed experiments.

(b) Children who experienced High Guidance in Training 1 (i.e., those in the High + High and High + Low conditions), were told that they would be evaluating experiments conducted by the experimenter and that they would be asked to evaluate the experiment on whether it was a 'smart' or 'not smart' way to test the target variable. The child observed while the experimenter set up a series of confounded and unconfounded experiments (two of each kind), and the child was asked whether the experiment was a 'smart' or 'not smart' way to test the target variable. Regardless of the child's response, the Experimenter then explained why each experiment was either smart or not smart, and provided the logical basis behind CVS. The child was then asked whether they could tell for sure from the experiment whether the target variable made a difference in how far the ball rolled, and were provided immediate feedback followed by an explanation about why one could or could not tell for sure whether the target variable was causal.

At the end of the Training 1 phase, a post-test was administered to children in all conditions. Children were asked to design an experiment to test each target variable in a procedure identical to the Ramps Pre-test. Scores ranged from 0 to 4.

## Phase 4: Training 2

This phase followed the identical procedures to the first training phase, the only difference being that children in the Low + High conditions now received a high degree of instructional guidance with the ramps apparatus and the children in the High + Low conditions now received low levels of instructional guidance. A post-test at the end Training 2 was scored identically to the post-test at the end of Training 1. The differences between the High and Low Guidance conditions are summarized in Table 1.

The remaining test phases (phases 5-10) consisted of a series of transfer tests that were meant to assess the robustness of children's ability to apply CVS to novel situations. In all transfer phases, no feedback was ever provided to children and the dependent variable was always how many experimental set ups children designed that were consistent with the logic of CVS.

Table 1: Similarities and differences between the two types of instruction

| | Amount of Instructional Guidance | |
|---|---|---|
| | High | Low |
| Goal Setting | "I'm going to design an experiment to test whether X makes a difference in how far the ball rolls." | "Can you design an experiment to test whether X makes a difference in how far the ball rolls?" |
| Number of Experiments | 4 | 8 |
| Design of Experiments | By Experimenter | By Child |
| Inquiry Questions | "Is this a smart or not smart experiment?" "Can we tell for sure from this experiment whether X made a difference?" | "Why did you set up your experiment that way?" "Can we tell for sure from this experiment whether X made a difference?" |
| Explanations | Experimenter explained why an experiment was smart or not smart, and why the child could or could not tell for sure whether X made a difference in the outcome | No Explanations |

### Phase 5: Springs Transfer Test

The materials for this phase consisted of a set of eight springs that varied across three different dimensions: length (long or short), spring width (wide or narrow), and coil width (thick or thin). A wooden centerpiece that children could hang the springs from was displayed on a table and two weights were present that could be attached to the springs in order to observe their stretching. Children were asked to design an experiment to test whether each of the variables (in a random sequence) made a difference in how far the spring stretched. Children tested all three variables and the child was assigned a 1 for CVS setups and a 0 for all other combinations. Scores ranged from 0 to 3.

### Phase 6: Car Design Transfer Test

A computer program (adapted from Klahr, Triona, & Williams, 2007) presented on a laptop computer allowed children to design cars to test what variables affected how far they traveled. Four variables – the body (long or short), the back axle (thick or thin), the front wheels (large thick, large thin, or small), and the back wheels (large thick, large thin, or small) – could affect how far the cars traveled. Children were told that they would be testing cars to figure out what made them travel farther, and that the experimenter would record the cars they made on a notepad to which they could refer back to if they forgot what cars they had constructed. Children proceeded to test each variable, which was indicated by the experimenter in a fixed sequence. In total, children had the opportunity to design 10 cars and the total possible chances of demonstrating CVS was 6. Therefore, scores ranged from 0 to 6.

### Phase 7: Ramps Transfer Test

The materials, procedure, and scoring in this phase were identical to the Ramps Pre-test.

### Phase 8: Story Transfer Test

The materials, procedure, and scoring in this phase were identical to the Story Pre-test.

### Phase 9: Remote Paper and Pencil Test

This phase took place roughly 5 months from training, when children were in the $4^{th}$ grade. Children sat at their regular classroom desks and completed a nine-question paper and pencil test that showed various experimental comparisons. Children were asked to identify which experiments were good and bad comparisons, and if the child identified an experiment as a bad test, they were asked to change the experiment to make it a good test. Three of the nine questions were good tests (CVS), whereas six questions were bad tests. Children were given a score of 1 if they correctly identified the test as either good or bad, and a score of 1 if they correctly changed the experiment. Scores ranged from 0 to 15.

### Phase 10: Remote Ramps Transfer test.

The materials, procedure, and scoring in this phase were identical to the Ramps Pre-test and Post-test, but took place 5 months after training.

## Results

Means and standard errors of the mean for each test phase by condition are displayed in Figure 1. A 4 (training condition) x 8 (test phase) mixed ANOVA on the first 8 test phases revealed a main effect of condition ($F(3, 38)$

= 6.21, $p < .005$), a main effect of test phase ($F(7, 266) = 43.64$, $p < .001$), and a significant interaction between them ($F(21, 266) = 2.05$, $p = .005$) (for the remainder of this report, statistics at each test phase will be reported for each group in the following order: 1) High + High, 2) High + Low, 3) Low + High, and 4) Low + Low) (story pre-test: $M1 = .23$, $SE1 = .06$, $M2 = .20$, $SE2 = .04$, $M3 = .28$, $SE3 = .05$, $M4 = .18$, $SE4 = .03$. For the ramps pre-test: $M1 = .25$, $SE1 = .08$, $M2 = .20$, $SE2 = .08$, $M3 = .10$, $SE3 = .06$, $M4 = .16$, $SE4 = .08$).

Post-hoc analyses (1) revealed no differences between any of the groups at the story pre-test (all $p$'s > .40) or the ramps pre-test (all $p$'s > .54). At the end of Training 1, however, the mean scores of the two groups receiving high levels of guidance improved dramatically from the ramps pre-test (all $p$'s < .001) while the groups receiving low amounts of guidance (i.e. Low + High and Low + Low) did not evidence reliable improvements (all $p$'s > .32). The groups receiving high guidance also performed at significantly superior levels when compared to the low guidance groups (all $p$'s < .01) (Training 1: $M1 = .98$, $SE1 = .03$, $M2 = .98$, $SE2 = .02$, $M3 = .50$, $SE3 = .14$, $M4 = .48$, $SE4 = .13$).

After Training 2, the Low + Low and the Low + High groups were no different from their performance after Training 1 or from their performance during the Ramps Pre-test (all $p$'s > .16), while the High + High group and the High + Low group continued to perform at ceiling levels (all one-tailed $p$'s > .21). Performance of both the High + High group and the High + Low group was significantly superior to the Low + Low group (all $p$'s < .05) (Training 2: $M1 = 1$, $SE1 = .00$, $M2 = .98$, $SE2 = .02$, $M3 = .73$, $SE3 = .14$, $M4 = .57$, $SE4 = .14$).

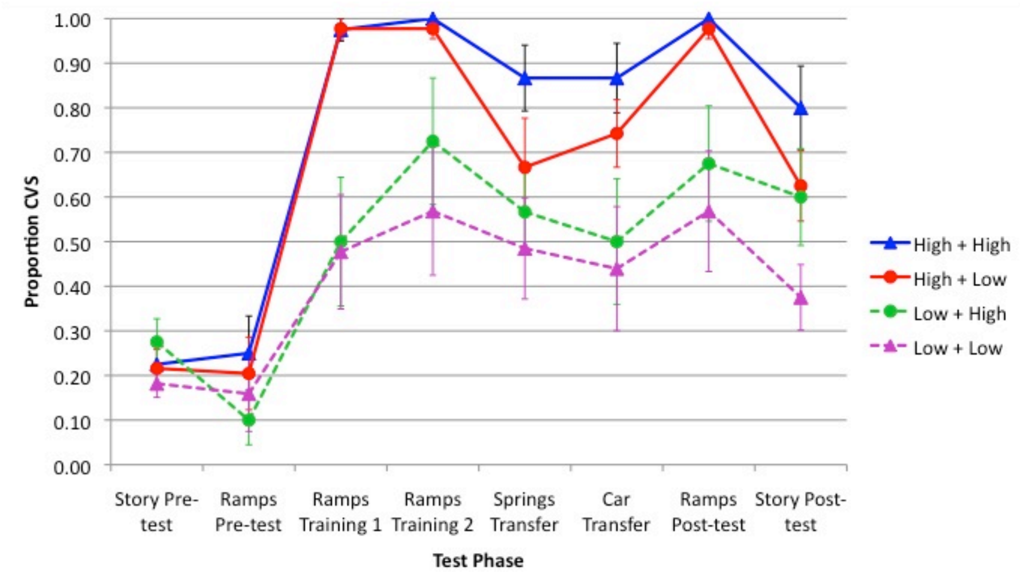



Figure 1. Means for each condition by the first 8 test phases: Error bars represent standard errors of the mean.

Of particular interest is the performance of children in the Low + High condition at the end of Training 2. Even though these children had just completed a training phase involving direct instruction, their performance was well below the ceiling levels of the High + Low group (one-tailed $p < .05$). To further examine the differences in learning gains between the High + Low and Low + High groups, a 2 (condition: High + Low vs Low + High) x 2 (phase: Ramps Pre-test vs Training 2) mixed ANOVA was conducted and this resulted in a significant interaction $F(1, 19) = 7.35$, $p = .01$, with the learning gains greater for the High + Low group. Thus, even though both groups were provided high *and* low amounts of instructional guidance during training, the order of these two types of instruction had substantial effects on learning, with the learning gains favoring the students in the High + Low group. However, since this finding might be due to low sample size, we would expect the effects to persist across a variety of transfer tests. This hypothesis was thus put to the test in the remaining transfer phases of the study.

The Springs Test was the first opportunity to demonstrate this transfer. Post-hoc analyses for this phase revealed statistically significant differences between the High + High and the Low + Low groups ($p < .05$) favoring the High + High group. All other comparisons were not statistically significant (all $p$'s > .17) (springs transfer: $M1 = .87$, $SE1 = .07$, $M2 = .67$, $SE2 = .11$, $M3 = .57$, $SE3 = .09$, $M4 = .48$, $SE4 = .11$). At the car design transfer test, significant differences were found between the High + High and the Low + Low groups ($p = .05$) with no other comparisons showing reliable differences (all $p$'s > .13) (car design transfer: $M1 = .87$, $SE1 = .08$, $M2 = .74$, $SE2 = .08$, $M3 = .50$, $SE3 = .14$, $M4 = .44$, $SE1 = .14$).

The High + High and Low + Low groups showed a significant difference at the Ramps Post-test ($p < .05$), as well as the High + Low and the Low + Low groups ($p < .05$). A marginally significant difference was found between the High + High and the Low + High groups ($p = .10$), with no other comparisons evidencing significance. Of interest was whether children performed at levels greater than Ramps Pre-test: here we found that all groups except the Low + Low groups showed a reliable improvement from the Ramps Pre-test to the Ramps Post-test (all $p$'s < .05) (ramps post-test: $M1 = 1$, $SE1 = .00$, $M2 = .98$, $SE2 = .02$, $M3 = .68$, $SE3 = .13$, $M4 = .57$, $SE1 = .14$).

The only reliable difference at phase 8 (Story Post-test) was between the High + High and the Low + Low groups ($p < .05$). No other comparisons were significant (all $p$'s > .32). We also compared children's post-test performance to their pre-test performance on the story problems and found that the High + High group evidenced a statistically significant improvement in performance ($p = .05$), the High + Low group showed a marginally significant improvement ($p = .07$), and the Low + High and Low + Low groups did not evidence any reliable improvements (all $p$'s > .16) (story post-test: $M1 = .80$, $SE1 = .09$, $M2 = .63$, $SE2 = .08$, $M3 = .60$, $SE3 = .11$, $M4 = .38$, $SE4 = .07$).

Five months after training, we assessed children's ability to apply CVS. Thirty-nine of 42 children were retained for these phases. ANOVA's for the remote paper and pencil tests evidenced no significant differences at either test phase (all $p$'s > .13) (5 month paper and pencil test: $M1 = .91$, $SE1 = .06$, $M2 = .66$, $SE2 = .09$, $M3 = .65$, $SE3 = .10$, $M4 = .61$, $SE4 = .12$. For the 5 month ramps test: $M1 = .97$, $SE1 = .03$, $M2 = .90$, $SE2 = .06$, $M3 = .77$, $SE3 = .13$, $M4 = .64$, $SE4 = .15$). However, because the null differences could be due to low sample sizes, we categorized students into CVS experts and non-experts. Experts on the remote paper and pencil test were defined as children who scored at least 13 of 15, and on the ramps post-test they were defined as children who designed at least 3 of 4 experiments as CVS. Distributions of experts and non-experts for each test are displayed in Figure 2. Chi-squared tests of independence revealed marginally significant differences in the distribution of experts at both the paper and pencil ($\chi^2(3, 39) = 7.12$, $p = .07$) and ramps tests ($\chi^2(3, 39) = 6.94$, $p = .07$). At the paper and pencil test, there was a greater proportion of experts in the High + High group compared to all other groups (all $p$'s < .05), while at the ramps test, all groups had greater proportions of experts than the Low + Low group (all $p$'s < .05).

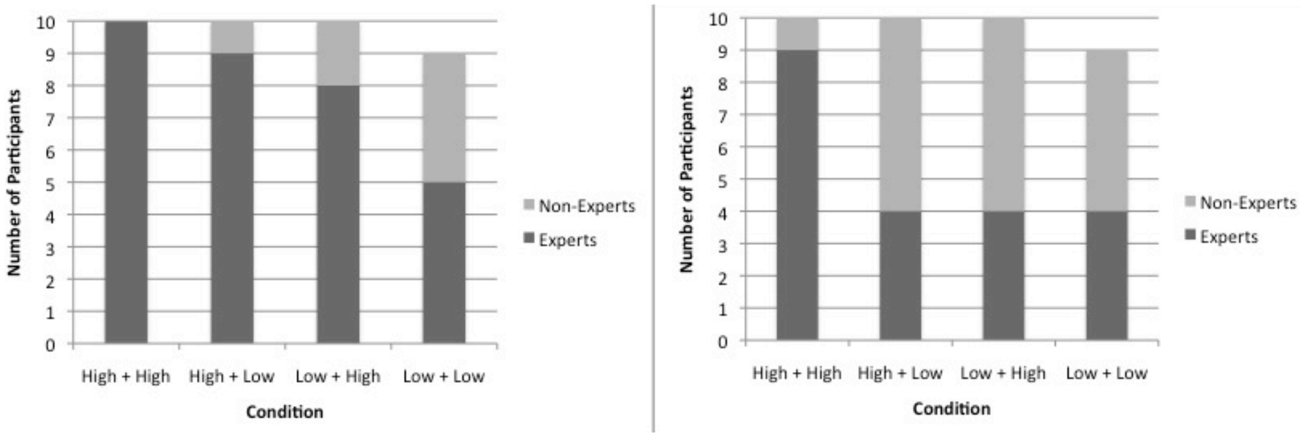



Figure 2. Left) Number of participants who were experts on the 5-month ramps transfer test by condition. Right) Number of participants who were experts at the 5-month paper and pencil transfer test by condition.

## Discussion

The most consistent finding of the present study was the superior performance of the High + High group when compared to the Low + Low group at every test phase after training. This suggests that the incorporation of direct instruction can be a powerful strategy that, coupled with inquiry teaching, is capable of promoting robust learning and transfer of scientific experimentation skill in children. Our findings also extend prior research on children's CVS learning, which has shown that a) minimally guided instruction often produces only nominal learning gains in children (Chen & Klahr, 1999; Klahr & Nigam, 2004) and b) the benefits of high amounts of guidance are not only immediate, but often continue to persist even after delayed periods of time (Chen & Klahr, 2004; Klahr & Nigam, 2004; Strand-Cary & Klahr, 2009; Klahr, Triona, & Williams, 2007).

A second goal of the present study was to examine what order of instructional guidance might prove most beneficial to student learning. While students' performance immediately after training suggested that the High + Low group may have benefited more from instruction than the Low + High group, we failed to find any differences between these groups at later points in transfer, suggesting that this initial effect after training may have been unique to that point in assessment. Moreover, in a subsequent follow-up study aimed to assess whether the differences between the High + Low and Low + High group would replicate, we divided a group of 30 third-graders ($M = 8.76$ years, $SD = .27$ years) evenly between the High + Low and Low + High groups and found no differences in learning across three separate transfer tests (averaged across all three tests, means were .69 and .63 for the High + Low and Low + High groups, respectively) independent-samples $t(28) = .15$, *ns*.

However, this follow-up study did replicate the finding that when students received high amounts of instructional guidance with ramps, they improved significantly from a ramps pre-test (for both groups, *p*'s < .05), while the Low + High group did not improve from the ramps pre-test after receiving low instructional guidance in the ramps domain (*p* = .35) (means for students ramps pre-test, ramps training 1 post-test, and ramps training 2 post-test were .14, .66, and .82 for the High + Low group, and .09, .13, and .72 for the Low + High group). Thus, it appears that, in the context of experimental design, young children learn just as well from different sequences of instructional guidance, as long as they are provided with high amounts of guidance at some point during learning.

The five-month transfer tests in the present study allowed us to assess how flexibly children could apply their CVS knowledge after extensive periods of time in 1) an identical domain to which they were trained (i.e. ramps), and 2) a domain that was removed from their original training (i.e. paper and pencil tests). Here, a large proportion of children in the High + High group evidenced flexible application of CVS to the paper and pencil post-test, whereas, only about half of the children demonstrated this proficiency in the other three groups. This result is striking when compared to the ramps transfer test, where a majority of children who had received direct instruction continued to perform at ceiling levels, and only about half of the children in the Low + Low group demonstrated this proficiency. It seems that, when transfer pertains only to time, children who receive high instructional guidance perform well on the materials on which they were trained, even after 5 months, and more so than children who do not receive high guidance. However, when transfer requires the application of knowledge over remote periods of both time *and* task, only students who received extensive amounts of direct instruction consistently applied CVS to novel situations beyond that of their training (i.e. the High + High group). Thus, results from this 5-month post-test strongly underscore the powerful effect that high levels of guidance can have on student learning.

Taken together, the present findings support the conclusion that students who are relative domain novices do not benefit more from any particular sequence of instruction, but that they do benefit – especially in the long run – from multiple lessons in which high amounts of guidance are provided. While this conclusion may, on the surface, seem conflicting with IPL and the concept of productive failures (2), it is important to note that studies reporting instructional sequence effects have often included other elements that seem key to acquiring differentiated knowledge, such as the use of contrasting cases, which we have left out of our instruction (Schwartz & Bransford, 1998). Students are also older in these studies, the youngest being 8th grade students, whereas we are investigating learning in 3rd grade students: There is little doubt that younger children differ from older children in many ways, including their working memory capacity, domain knowledge, and recruitment of metacognitive strategies. The disparity between our findings, however, prompts discussion of the circumstances that may be crucial to IPL and productive failures. One fundamental prerequisite for a failure to serve as a productive learning experience seems to be that the learner must first realize that failed solutions are indeed, failures. Without such knowledge, students may continue to practice and strengthen the use of incorrect strategies, therefore failing to acknowledge the need for more efficient solutions offered through instruction. Alternatively, if the student is aware that his/her strategies are suboptimal, the student may be more prone to recognize the value of instruction. Indeed, students in Schwartz and Martin (2004) could readily observe whether their formulas for variance accounted for the data they were trying to describe. Kapur (in press) also showed that students in less structured settings often exhibited lower confidence in their solutions, suggesting they were aware of the shortcomings of these solutions. In contrast, many of the students who designed experiments in our study gave little indication that they were aware that their strategies were incorrect, frequently responding that they *could* tell for sure whether the target variable made a difference in the outcome, even when their experiment was confounded by multiple factors. While the interactions between characteristics of individual learners, the learning domain, and instructional support are no doubt a complex issue and beyond the scope of the present study, we propose that attempting to understand the relative contributions of each of these factors, as well as how they interact, would be a productive area for future research to move towards.

What this study does contribute to our thinking about instructional design is that it refutes widely held beliefs that direct instruction is insufficient for promoting robust learning (Dean & Kuhn, 2007) and that shallow transfer is one of the hallmarks of this form of instruction. On the contrary, the present study suggests that, for domain novices, minimally guided instruction that neglects direct instruction may miss opportunities to optimize student learning. We recognize that over the course of a students' life, he/she may experience a great number of instructional events, and that this study addresses only the very earliest of these instructional experiences. We thus agree with, and emphasize that a combination of various instructional approaches may indeed prove of large value to a learner throughout the course of their education (Koedinger & Aleven, 2007). At least in early stages of learning, however, the present study suggests that providing high amounts of guidance, particularly in the form of multiple phases of direct instruction, can be a useful strategy in promoting robust understanding of scientific experimentation skill.

## Endnotes

(1) All between subject post-hoc tests are Tukey adjusted, and within subject post-hoc tests are Bonferonni adjusted.
(2) We purposefully exclude discussion of the Expertise Reversal effect here, as this view could explain the present findings by positing that, despite receiving training, children were relative domain novices at this stage in learning.

## References

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*, 1098-1120.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91,* 384-397.

Germann, P.J., Aram, R., & Burke, G. (1996). Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *Journal of Research in Science Teaching, 33*, 79–99.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509-539.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93,* 579-588.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558-568.

Kapur, M. (2008). Productive Failure. *Cognition and Instruction, 26*, 379-424.

Kapur, M. (In press). Productive failure in mathematical problem solving. *Instructional Science*

Kirschner, P.A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75-86.

Klahr, D. (2009) "To every thing there is a season, and a time to every purpose under the heavens": What about Direct Instruction? In S. Tobias and T. M. Duffy (Eds.) *Constructivist Theory Applied to Instruction: Success or Failure?* Taylor and Francis.

Klahr, D. & Li, J. (2005) Cognitive Research and Elementary Science Instruction: From the Laboratory, to the Classroom, and Back. *Journal of Science Education and Technology*, 4, 217-238.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*, 661-667.

Klahr, D., Triona, L.M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*,

Koedinger, K.R. & Aleven, V. (2007). Addressing the Assistance Dilemma in Experiments with Cognitive Tutors. *Educational Psychology Review, 19,* 239-264.

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about controlling variables? *Psychological Science, 16,* 866-870.

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist, 42,* 109-113.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*

McDaniel, M.A., & Schlager, M.S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction, 1990, 7,* 129-159.

National Research Council (1996). The National Science Education Standards. Washington DC: National Academy Press.

Schwartz, D.L., & Bransford, J. (2006). A Time for Telling. *Cognition and Instruction, 16,* 475-522.

Schwartz, D.L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cogniton and Instruction, 22*, 129-184.

Strand-Cary, M., & Klahr, D. (2009). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development, 23,* 488-511.

Siegler, R., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46*, 606 - 620.

Toth, E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction, 18,* 423-459.

## Acknowledgments

We thank Howard Seltman, Anna Fisher, Marsha Lovett, Bob Siegler, Stephanie Siler, Jamie Jirout, Kevin Willows, Audrey Russo and Cressida Maguro for their various intellectual contributions. This work was supported by the Institute of Education Sciences under grant #R305B040063 to Carnegie Mellon University.

# Comparing Pedagogical Approaches for the Acquisition and Long-Term Robustness of the Control of Variables Strategy

Michael A. Sao Pedro, Janice D. Gobert, Juelaila J. Raziuddin
Worcester Polytechnic Institute
100 Institute Rd. Worcester, MA 01609
Email: mikesp@wpi.edu, jgobert@wpi.edu, juelaila@wpi.edu

**Abstract:** This study compared three pedagogical approaches on the acquisition and robust understanding of the control of variables strategy (CVS). In Sao Pedro et al. (2009), we showed that two direct learning conditions (with and without reification) significantly outperformed the discovery learning condition for constructing unconfounded experiments starting from an initially multiply confounded experimental setup. In the study described here, we retested 57 students six months later on constructing unconfounded experiments in a virtual ramp environment, solving problems requiring CVS, and explaining CVS. Collapsing over time, we found that the direct+reify condition had more robust learning than either the direct-no reify or discovery learning conditions on constructing unconfounded experiments. At the delayed posttest, we found a strong trend favoring the direct+reify condition over the other conditions as measured by tasks designing unconfounded experiments starting from a multiply confounded state.

## Introduction

Currently, there is a debate in the science education community regarding the effectiveness of discovery vs. direct instruction. Critics of discovery learning claim that it may be less effective when compared to instructional approaches with emphasis on guidance (Kirschner, Sweller, & Clark, 2006). In particular, it has been found that during open-ended inquiry, students can have many false starts (Schauble, 1990), and have difficulty designing effective experiments, forming testable hypotheses, adequately monitoring what they do (de Jong, 2006), linking hypotheses and data, and drawing correct conclusions (Klahr & Dunbar, 1988; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). On the other hand, direct instruction, in its traditional form, has also received criticism, namely, that it can turn into rote instruction leading students to lose engagement and develop inert knowledge that cannot be flexibly applied or transferred (Bereiter & Scardamalia, 1989).

Within these approaches, researchers have begun to determine if these instructional paradigms yield successful transfer of knowledge and skills. Some have shown that explicitly teaching *strategies and skills* in one context can successfully transfer to another context (Klahr & Nigam 2004; VanLehn et al., 2005) or domain (Chi & VanLehn, 2007). More specifically and relevant to the present study, Klahr and Nigam (2004) found that students who were taught the control of variables strategy (CVS) in a direct learning condition significantly outperformed those in a discovery learning condition on a near-transfer test of CVS. Furthermore, those who mastered CVS, irrespective of learning condition, outperformed non-masters on a near-transfer test of this skill. These results suggest that the purported benefits of discovery learning, particularly the deeper learning, may not always occur. These Klahr and Nigam (2004) findings, though, are not without critique (cf., Hmelo-Silver, Duncan, & Chinn, 2007). Others such as Kuhn (2005) have criticized Klahr and Nigam because they did not test students' knowledge about when and why to use CVS.

In a similar vein, others have explored the degree to which these instructional paradigms lead to long-term robustness of skills. For example, Strand-Cary and Klahr (2008) compared the long-term effectiveness of guided and unguided approaches to teaching CVS by repeatedly testing students over a much longer period of time, up to 3 years after their initial intervention. They found significant differences in skill at constructing unconfounded experiments favoring their explicit instruction condition immediately after their intervention, but that 3 months later, those in their exploration condition caught up to the explicit instruction group on average. Furthermore, those who mastered CVS by the 3 month mark, irrespective of condition, significantly outperformed non-CVS masters on transfer tasks 3 years later. Dean and Kuhn (2006) explored the role of practice and engagement within the direct and discovery frameworks on robust acquisition of CVS. They found that direct instruction alone did not lead to robust acquisition; practice and engagement, irrespective of initially receiving instruction, produced deep, lasting learning. In our work, we explored the role that self-explanation played in long-term retention and transfer of CVS skill.

Our first study, Sao Pedro, Gobert, Heffernan, & Beck (2009), an extension to Klahr and Nigam's (2004) study, compared the effectiveness of two types of direct instruction versus discovery with regard to the acquisition and transfer of CVS. We hypothesized that the self-explanation component in Klahr and Nigam's direct instruction condition could have played a role in students' acquisition of CVS. By adding our third condition, direct no-reify, which removed prompting for student explanations, we empirically tested if the explanations in the direct+reify condition affected the acquisition of CVS. The terms "direct" and "discovery" have slightly different meanings than is reflected in our learning conditions; that is, these terms typically represent polar opposites in terms of level of *directedness* given to students. Our direct instruction conditions portray variants of guided inquiry in which students are taught the procedure of CVS in a concrete context, a virtual ramp environment. Our results showed that in an immediate posttest following the interventions, students in both direct conditions significantly outperformed students in the discovery condition on tasks involving creating unconfounded experiments starting from an initally multiply confounded setup. The direct conditions, however, did not significantly differ from each other.

In the present study, we tested the efficacy of the reification task (self-explanation) on the acquisition and robust understanding of CVS over time, 6 months after the original Sao Pedro et al. (2009) study by administering a delayed posttest on participants from our original study. We compared the conditions on the understanding of CVS as evidenced by their skills at constructing unconfounded ramp setups, explaining the CVS procedure, and solving multiple choice and open response problems requiring CVS. Self-explanation has been found to support deep learning (Chi M. T., 1996) and knowledge intgegration (Linn & Hsi, 2000). Thus, in this study, we hypothesized participants in the direct+reify group would outperform the other conditions even though the direct-no reify group showed the same performance as direct+reify for certain items at the immediate posttest (Sao Pedro et al, 2009).

## Method

### Participants

Participants were seventh-grade students from a public middle school in central Massachusetts. We chose this group because middle school may be the time to optimally learn model-based inquiry skills (Schunn, Raghavan, & Cho, 2007). Participants all had the same science teacher through the school year who taught five heterogeneously grouped sections. Students at this school typically struggle with science. For example, in 2008, 92% of eighth grade students at this school scored below proficient on the science MCAS exam, a standardized test administered by the state (Massachusetts Department of Elementary and Secondary Education (ESE), 2008).

All class sections participated in the experiment. For this paper, we concerned ourselves only with the performance of those students who completed our initial intervention and posttest (Sao Pedro et al, 2009) as well as the delayed posttest. Further, we did not include data from students on individual education programs (IEPs) and one additional student who used an incorrect web browser during the initial intervention, leaving a total of 57 students in the analyses for this study.

### Materials

We used The Science ASSISTment System (Gobert et al., 2007; Gobert et al., 2009), a web-based intelligent tutoring system, to host our materials, run randomized controlled experiments and log time-stamped student interactions with the system. We used several types of activities to assess students' acquisition of CVS. Students practiced authentic inquiry by constructing unconfounded experiments using our virtual ramp environment. The same ramp activity acted as a near-transfer performance assessment. We also developed a far transfer test involving several multiple choice and open response questions. Some were designed by us and others were obtained from a study conducted by Strand-Cary and Klahr (2008).

#### Ramp Environment

The ramp environment (Figure 1) was developed using the OpenLaszlo framework (www.openlaszlo.org). We created different kinds of questions by embedding the ramp environment within the ASSISTment System. The ramp apparatus has four variables that can be manipulated: *surface* (smooth or rough), *ball type* (golf or rubber), *steepness* (low or high), and *run length* (long or short). The objective is to set up the ramps so that the target variable is contrasted and all other variables are held constant. Pressing the "run" button causes each ball to roll down its respective ramp. Depending on each ramp's settings, the ball will roll different distances. Participants submit their final setup using the "submit" button. Pressing "reset" causes the balls to be placed back on the ramp and clears the distances rolled. On each press of the "run" or "submit" button, student information is logged. This includes a timestamp of the run or submission, the correctness of their setup in terms of CVS, and the current and previous ramp value settings.

Unlike the physical and virtual environments used in Klahr and Nigam (2004) and Strand-Cary and Klahr (2008) in which participants set up ramp conditions starting from a blank slate, our ramp setup was always initially set. There was always an initial condition students must change in order to create an unconfounded experiment. In a sense, this activity can be viewed as an experimental setup repair task since students must morph a given setup into one that follows CVS. The experiment's starting state could initially be *unconfounded* (all variables are controlled), *singly confounded* (one variable is not controlled), *multiply* *confounded* (more than one variable is not controlled), and/or *uncontrasted* (the target variable is unchanged). Our instruction and feedback format was also different from previous studies; the CVS explanations given to the direct conditions were entirely computer-based whereas the other studies both used human tutoring.

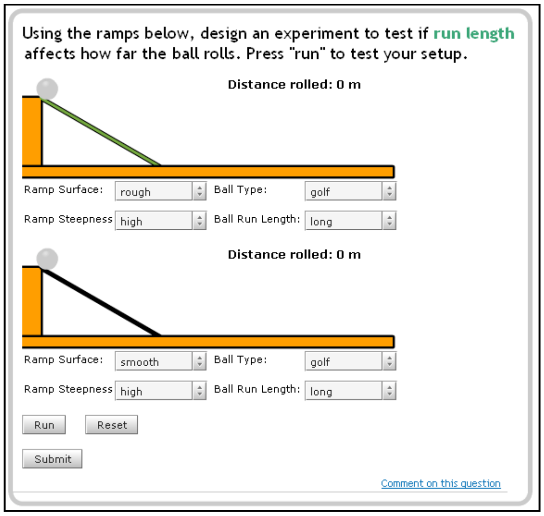



Figure 1. ASSISTment question using the ramp environment. This setup is uncontrasted and singly confounded because the target variable, run length, is the same for each setup and one extraneous variable, surface, is not controlled.

## Multiple Choice and Open Response Transfer Items

In Sao Pedro et al (2009), we developed three multiple choice and two open response items to assess students' skills in using CVS to solve problems and in communicating the procedure. We reused these exact items again for this study. The multiple choice items, referred to as the WPI CVS Inquiry items, asked students to identify an unconfounded experiment, identify a CVS procedure step, and determine an appropriate experimental setup with a valid control condition. The two open response items asked students to describe in their own words the CVS procedure. Specifically, they asked how (1) they "could determine if one particular variable affects how far the ball rolls in the ramp experiment" and (2) "when there are many variables that can be changes, explain how [they] could determine how each variable affects the distance the ball rolls." Students needed to clearly state that CVS was a procedure one could systematically repeat to find out how each variable affected the outcome.

For this study, we also added two multiple choice items and one open response item from a transfer test developed by Strand-Cary and Klahr (2008). Their multiple choice items on plant and seed growth required students to choose the correct experimental setup that contrasted a variable of interest and kept all others constant. Their open response item asked, without specific prompting for CVS usage, for students to critique a confounded experimental design that tested if beetles preferred to live in sun or shade.

## Procedure

This study determined the long-term effects of the differing interventions on CVS acquisition using both the original posttest materials and the new Strand-Cary and Klahr (2008) assessments. In our original October 2008 experiment, students were first pretested on their skillfulness at creating unconfounded experiments in the ramp environment without receiving feedback. We also tested them on the WPI CVS Inquiry items, multiple choice questions involving CVS. The ASSISTment System then randomly assigned each student to either the direct+reify, direct-no reify, or the discovery condition. In each condition, students practiced how to perform CVS in the context of the ramp environment. Following the intervention, students were tested on their skill at constructing unconfounded ramp setups, answering the same WPI CVS multiple choice questions as the pretest, and explaining in their own words the CVS procedure within the context of the ramp environment.

In both direct conditions, students were first asked to read an overview of CVS with examples of confounded and unconfounded ramp setups. In this overview, we did not explain that CVS could be used to solve problems in other domains. After reading the overview, students addressed whether a series of ramp setups had correctly controlled for variables. In the *direct+reify* condition, participants first responded if a given ramp configuration enabled them "tell for sure" if a variable affected how far a ball would roll by responding "yes" or "no". Next, they answered an open response question asking them to explain their reasoning. For the same ramp setup, students then were allowed to run the experiment as many times as desired *without changing* *the setups* and explained again if they could tell for sure that target variable affected how far the ball would roll. Finally, students read an explanation why the experiment was confounded or unconfounded for the target variable. If the setup was confounded, students were told which variables were confounded. Students in the *direct-no reify* condition followed the exact same procedure, except they were not asked the two open response questions. Students in the *discovery* condition were instructed to create experiments that tested if a particular variable affected how far the ball rolled. The discovery condition students were neither given the initial CVS explanation nor any feedback about the correctness of their experimental setups. All conditions practiced on six identical initial ramp configurations. For more details on the original experimental procedure, see Sao Pedro, Gobert, Heffernan, & Beck, (2009).

The present study took place six months later in June 2009 and acted as a delayed posttest of CVS skill. Throughout the school year, students used the ASSISTment system extensively for math class but not for science-related activities. All students in each class section took the delayed posttest irrespective of having participated in the original experiment, though we only analyze those students for whom we have posttest data from our October 2008 experiment. Students first answered the WPI CVS Inquiry multiple-choice questions developed in the previous study. Next, students answered the subset of Strand-Cary and Klahr (2008) questions, including one open response question. Students were then presented a reintroduction to the ramp environment that described the ramp variables and how to interact with the simulation. Students then constructed four unconfounded ramp setups with different target variables. The initial setups were identical to those in the previous study's posttest. Finally, students again answered the open response items in which they explained CVS in the context of the ramp experiment.

### Scoring

Multiple-choice questions were automatically scored by the ASSISTment system with a 1 if correct or 0 if incorrect. Correct ramp setups demonstrating CVS for the given target variable were awarded 1 point, 0 otherwise. All open response items were hand scored by two different graders according to a rubric. The Strand-Cary and Klahr (2008) open response question was graded out of 3 points, one point for identifying that the experimental setup was incorrect, and up to 2 points for correctly explaining the experiment's design flaws. Maximum points were only awarded if the student addressed the lack of control in the question's experimental design.

The two ramp open response items were combined into one measure to capture skill in correctly describing the CVS procedure, marked out of 3 points. Higher scores indicated deeper understanding of CVS. One point was given for identifying independent and dependent variables and the presence of some relationship between them. Two points were awarded for stating that the target variable needed to be contrasted and all other variables should be the same. Three points were given for understanding that CVS should be repeated for each variable to determine how each individually affected the outcome. For example, a student who received full credit answered as follows: "For each variable keep the others the same and change one variable at a time." When reviewing student responses, we realized a large number of students described how each specific variable affected the distance rolled instead of describing a procedure. If students answered with such evidential knowledge, they were scored according to a different rubric and were treated as missing data for this metric. Students falling into neither category, such as those responding "I don't know" received 0 points.

## Results

We analyzed 57 students' immediate and delayed posttest scores to determine which condition(s) yielded better performance on each CVS measurement: problem solving using CVS via the multiple choice items, articulation and explanation of CVS via the ramp open response items, and authentic inquiry performance with the ramp environment. In particular, we determine if self-explanation in the direct+reify condition led to better performance over time for the various near-transfer and far-transfer CVS tasks. We also analyzed if the findings favoring the direct+reify and direct-no reify groups over the discovery group on multiply confounded ramp items were robust for the delayed posttest. For each analysis of variance below, we report partial $\eta^2$ values as measurements for effect size.

### Differences between Groups on Transfer Items

In these analyses we looked for trend differences between groups at immediate and delayed posttest for the WPI CVS inquiry test and the ramp open response items. We also compared groups' performance on the Strand-Cary and Klahr (2008) items. Means and standard deviations for all transfer items are presented in Table 1. Only students who completed all the questions for a measurement were included in that measurement's analysis. We conducted a repeated measures ANCOVA to compare performance on the WPI CVS Inquiry multiple choice items over time. The WPI CVS Inquiry pretest was used as a covariate. Though time was not a significant within-subjects factor (Wilks $\lambda$=1.00, $F(1,48)=0.19$, $p$=.668, partial $\eta^2$=.004), the interaction effect between time and condition was marginally significant, Wilks $\lambda$=.907, $F(2,48)=2.47$, $p$=.095, partial $\eta^2$=.093, suggesting

Table 1: Means and standard deviations for the transfer items at pretest (t0), immediate posttest (t1), and delayed posttest (t2).

| | | *Direct+Reify* | | | *Direct-No Reify* | | | *Discovery* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Max | *M* | *SD* | *N* | *M* | *SD* | *N* | *M* | *SD* | *N* |
| WPI CVS MultCh t0 | 3 | 1.38 | 1.07 | 21 | 1.00 | 0.85 | 15 | 1.06 | 1.00 | 16 |
| WPI CVS MultCh t1 | 3 | 1.81 | 0.91 | 22 | 1.25 | 1.00 | 16 | 1.52 | 1.07 | 19 |
| WPI CVS MultCh t2 | 3 | 1.36 | 1.00 | 22 | 1.25 | 0.93 | 16 | 1.74 | 0.93 | 19 |
| Ramp CVS OpenR t1* | 3 | 1.06 | 1.11 | 18 | 1.57 | 0.79 | 7 | 1.08 | 1.12 | 13 |
| Ramp CVS OpenR t2* | 3 | 1.11 | 1.08 | 18 | 0.80 | 0.92 | 10 | 0.80 | 0.79 | 10 |
| Strand-Cary & Klahr t2 | 3 | 1.38 | 1.00 | 21 | 1.38 | 0.74 | 16 | 1.61 | 0.96 | 18 |

* Open response items include only those who responded with procedures, not evidential statements.

there may have been performance differences between the conditions as time progressed. As Table 1 indicates, performance in the direct-no reify condition was the same for the each posttest, *M*=1.25. Students in the direct+reify condition showed better performance on average at the immediate posttest, but did not maintain those scores. The discovery condition's performance increased on average as time progressed. Condition was not a significant between-subjects factor, $F(2,48)=1.78$, $p=.179$, partial $\eta^2=.069$. This indicated no performance differences between the groups when collapsing over time.

Another repeated measures ANCOVA determined if any intervention affected students' skill in explaining CVS. In this analysis, we compared only students who explicated procedures for constructing unconfounded experiments in response to the two ramp open response questions since some students explained which variables affected how the ball rolled instead of explaining a procedure. There was neither a significant effect for time (Wilks $\lambda=.98$, $F(1,21)=0.49$, $p=.492$, partial $\eta^2=.023$) nor for the interaction between time and condition, Wilks $\lambda=.93$, $F(2,21)=0.81$, $p=.460$, partial $\eta^2=.071$. Condition was also not a significant between-subjects factor, $F(2,21)=0.23$, $p=.800$, partial $\eta^2=.021$.

Finally, we analyzed if there were differences between the groups on the Strand-Cary and Klahr (2008) items using an ANOVA. We normalized the beetles open response item in order to equally weight all items for the dependent measure. No significant difference was found between the groups on this measure, $F(2,52) = 0.39$, $p=.681$, partial $\eta^2=.015$.

Taking these transfer results as a whole, we saw no significant differences in trends and no significant differences between conditions for any of the transfer item subsets. This suggests that our different learning conditions did not enable transfer from constructing unconfounded ramp experiments to solving more complex reasoning problems requiring CVS. As for explaining CVS, we believe we would expect to see differences between learning conditions if the questions were worded such that students answered with procedures rather than evidential statements. We believe this because there were significant differences between groups' overall skills in constructing unconfounded experiments in the ramp environment as discussed next.

## Comparison of Ramp Items

We compared each condition's skill in constructing unconfounded ramp experiments over time using a repeated measures ANCOVA. Total ramp scores at each time were dependent measures and total ramp pretest score was used as a covariate. Means and standard errors for ramp performance over time are shown in Figure 2. Four students did not take the original ramp pretest and were not included in this analysis, leaving 53 students. Within-subjects tests revealed no significant main effect for time, Wilks $\lambda=.97$, $F(1,49)=1.36$, $p=.249$, partial $\eta^2=.027$, and no significant interaction between time and condition, Wilks $\lambda=.92$, $F(2,49)=2.03$, $p=.143$, partial $\eta^2=.076$. However, condition was a significant between-subjects factor, $F(2,49)=3.26$, $p=.047$, partial $\eta^2=.117$, controlling for ramp pretest score. Post hoc comparisons revealed that the direct+reify condition constructed significantly more experiments adhering to CVS than the discovery condition (*M*=0.92, *SE*=0.43, *p*=.037, 95% *CI*=[0.06, 1.79]) and the direct-no reify condition (*M*=0.98, *SE*=0.45, *p*=.033, 95% *CI*=[0.84, 1.88]). No significant difference was found between the direct-no reify and discovery conditions, *p*=.903.

Since there was a significant between-subjects effect on condition, we also analyzed group differences solely at the time-2 delayed posttest using an ANCOVA with ramp pretest score as a covariate. The analysis yielded a main effect on condition, $F(2,49)=3.36$, $p=.043$, partial $\eta^2=.121$. The ramp pretest was not a significant covariate, $F(1,49)=1.38$, $p=.247$. Post hoc tests revealed that the direct+reify condition constructed significantly more unconfounded experiments in the delayed posttest as compared to the direct-no reify condition (*M*=1.34, *SE*=0.53, *p*=.014, 95% *CI*=[0.29, 2.40]). No significant differences between direct+reify and discovery, *p*=.131, and direct-no reify and discovery, *p*=.306 were found. As shown in Figure 2, the direct+reify condition maintained its higher performance at the immediate and delayed posttest. Though the direct-no reify condition improved at the immediate posttest compared to the pretest, the improvement is lost 6 months later at the delayed posttest. Also, the discovery condition's skills on this CVS authentic inquiry task increased as time progressed.

We also compared the number of students proficient at CVS (those who scored at least 3 out of 4 on the ramp) in each condition at the delayed posttest. We used only those who did not score a perfect 4 out of 4 on the ramp pretest (49 participants). Of those remaining, 59% (10 out of 17) in the direct+reify condition, 24% (4 out of 17) in the direct-no reify condition, and 33% (5 out of 15) in the discovery condition were proficient. Though the tallies favor direct+reify over the other conditions, the differences were not significant, $\chi^2(2)=4.43$, $p=.109$. We note that in the delayed posttest, there were more than twice as many CVS-proficient students in the direct+reify condition as compared to the direct-no reify condition.

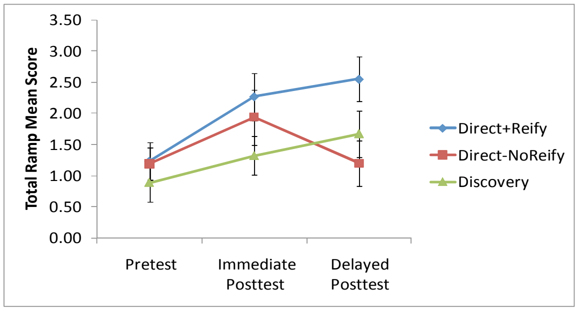



Figure 2. Means and standard errors for total ramp score by condition. The maximum score on this measure is 4.

## Comparison of Ramp Items by Level of Complexity

As in our previous 2009 study, we examined performance changes over time for ramp items of differing complexity, namely singly and multiply confounded initial setups. To determine time and condition effects we conducted a repeated measures MANCOVA using singly and multiply confounded ramp items at each time point as within subjects factors, condition as a between subjects factor, and singly confounded ramp pretest score as a covariate. This design mimics the design used in our 2009 analysis. There were neither main effects on the dependent variate for time ($p=.646$) nor for the time and condition interaction, $p = .31$. However, there was a significant between subjects effect on the dependent variate for condition, Wilks $\lambda=.73$, $F(4,96)=4.18$, $p=.004$, partial $\eta^2=.148$.

ANCOVAs on each dependent measure revealed a main effect for condition for multiply confounded ramp items, $F(2,49)=7.58$, $p=.001$, partial $\eta^2=.236$ but not for singly confounded ramp items, $F(2,49)=1.18$, $p=.156$, partial $\eta^2=.073$. Post hoc comparisons between conditions for multiply confounded items revealed that the direct+reify condition created significantly more unconfounded experiments starting from a multiply confounded state over both the discovery ($M=.62$, $SE=.19$, $p=.002$, 95% $CI=[0.25, 1.00]$) and direct-no reify conditions ($M=.64$, $SE=.20$, $p=.002$, 95% $CI=[0.24, 1.03]$). Again, no significant differences were found between direct-no reify and discovery, $p=.949$. As shown in Figures 3 and 4, the direct+reify condition maintained its performance edge over time as compared to the other two conditions. Also, though the direct+reify and direct-no reify conditions were similar at immediate posttest, the direct-no reify condition did not show long-term skill in constructing unconfounded experiments as shown by the drop in performance at the delayed posttest. Again, the discovery condition continued to increase in average performance over time.

Group differences for the time-2 delayed posttest were also analyzed using a MANCOVA with singly confounded ramp pretest items as a covariate. The dependent variate comprised of singly and multiply confounded ramp items at time 2 was marginally affected by condition, Wilks $\lambda=.84$, $F(4,96)=2.21$, $p=.074$, partial $\eta^2=.084$. The covariate was also marginally significant, Wilks $\lambda=.91$, $F(2,48)=2.41$, $p=.100$, partial $\eta^2=.091$. Observing the differences more closely with ANCOVAs on each dependent measure revealed a significant effect for condition for multiply confounded ramp items, $F(2,49)=4.23$, $p=.020$, partial $\eta^2=.147$ and a marginally significant effect for condition for singly confounded ramp items, $F(2,49)=2.82$, $p=.069$, partial $\eta^2=.103$. Pairwise comparisons revealed a significant difference in constructing experiments that followed CVS starting from a multiply confounded state favoring direct+reify over direct-no reify ($M=.73$, $SE=.26$, $p=.007$, 95% $CI=[-0.21, 1.25]$), and a marginally significant difference favoring direct+reify over discovery ($M=.48$, $SE=.25$, $p=.061$, 95% $CI=[-.02, 0.97]$). The direct-no reify condition was not significantly different than discovery learning condition, $p=.350$.

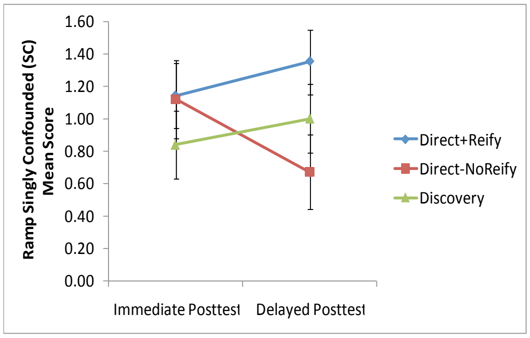



Figure 3. Means and standard errors for singly confounded ramp items (max=2). The pretest was omitted since it had only 1 singly confounded item.

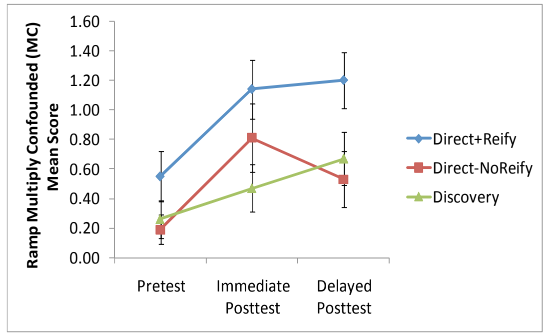



Figure 4. Means and standard errors for multiply confounded ramp items (max=2).

## Discussion and Conclusions

The primary goal of this research was to compare three pedagogical approaches in terms of their efficacy at fostering robust understanding of CVS as measured by successful application of CVS and long-term retention of this skill. Specifically, we examined if direct instruction coupled with self-explanation (direct+reify) afforded any long-term benefits on CVS acquisition over direct instruction without self-explanation (direct-no reify) and discovery learning. In Sao Pedro, et al. (2009), we found that students in each learning condition did not perform significantly better than each other on CVS problem solving and overall construction of unconfounded experiments. However, those in either direct instruction condition were better at correcting the more difficult multiply confounded setups to correctly contrast variables than the discovery condition at the immediate posttest. When incorporating the results of the delayed posttest administered six months later, direct instruction with explanation yieded a deeper understanding of CVS, collapsing over time. Looking specifically at the delayed posttest, those in the direct+reify condition significantly outperformed the direct-no reify condition on authentic inquiry items starting from a multiply confounded state. Additionally, those in the direct+reify condition outperformed those in the discovery condition, but these results were not strong enough to reach statistical significance. The instruction method did not impact long-term skill to articulate CVS nor CVS problem solving skill. The differences between our direct and discovery conditions are not as drastic as those of Klahr and Nigam (2004); this may be attributed to our choice to use text as the means to communicate instruction as opposed to audio as Klahr and Nigam did.

Our results are comparable to both Strand-Cary and Klahr (2008) and Dean Jr. and Kuhn (2006). Like us, Strand-Cary and Klahr reported that in a delayed posttest, students in their “discovery” condition, on average, rose to the levels of their explicit instruction condition at constructing unconfounded experiments. We concur with their interpretation that students’ engagement at constructing unconfounded experiments, though initially unsuccessful, may have made them more aware of the necessity to control for variables. Furthermore, it may be possible that students’ initial failure primed them for future encounters with the concept (Kapur, 2008). Further research needs to be conducted to empirically test these claims. Nonetheless, our collective findings suggest that it is possible for unguided exploration to reach levels comparable to explicit instruction with self-explanation, albeit more slowly. Additionally, Dean Jr. and Kuhn (2006) found that participants who practiced CVS regularly, irrespective of whether they received direct instruction or not, were better able to apply the strategy and make inferences when compared to those receiving direct instruction and not practicing. In their study, the practice condition dyads provided explanations to each other and entered them into the computer; these activities could have significantly affected students’ knowledge of CVS. In other words, it is difficult to ascertain whether the robustness of their findings is due to practicing the skill or providing the explanation.

In this paper, we have shown that the reification process (self-explanation), when added to direct instruction of the skill, supports both efficient acquisition and long-term retention of CVS, as measured by authentic inquiry performance. Of interest, we also found that on the same measurement, a nontrivial number of students in our unguided discovery condition were also able to perform, on average, at levels comparable to the direct+reify condition 6 months after the initial intervention. To address which behaviors within each intervention led to robust acquisition of CVS, we are looking at the quality of self-explanations, student responses, and all log files of student interactions within the ramp microworld. This will allow us to characterize the affordances of each condition for learning CVS and examine interactions with student variables such as prior knowledge of inquiry, in particular, knowledge of designing and conducting experiments. A fuller description such as this will provide important data towards our eventual goal of developing adaptive scaffolding support for CVS as it is a necessary but not sufficient skill for conducting inquiry (Gobert et al., 2007; Gobert et al., 2009).

## References

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L.B. Resnick (Ed.). In L. R. (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T. (1996). Constructing Self-Explanations and Scaffolded Explanations in Tutoring. *Applied Cognitive Psychology, Vol. 10, S33-49.*

Chi, M., & VanLehn, K. (2007). Accelerated Future Learning via Explicit Instruction of a Problem Solving Strategy. *13th International Conference on Artificial Intelligence in Education* (pp. 409-416). Amsterdam: IOS Press.

de Jong, T. (2006). Computer Simulations - Technological advances in inquiry learning. *Science, 312* , 532-533.

Dean Jr., D., & Kuhn, D. (2006). Direct Instruction vs. Discovery: The Long View. *Science Education* , 384-397.

Gobert, Janice (Principal Investigator); Heffernan, Neil; Koedinger, Ken; Beck, Joseph (Co-Principal Investigators). (2009). ASSISTments Meets Science Learning (AMSL; R305A090170). Awarded February 1, 2009 from the U.S. Dept. of Education.

Gobert, Janice (Principal Investigator); Heffernan, Neil; Ruiz, Carolina; Kim, Ryung (Co-Principal Investigators). (2007). AMI: ASSISTments Meets Inquiry (NSF-DRL# 0733286). Awarded September 2007 from the National Science Foundation.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Krischner, Sweller, and Clark (2006). *Educational Psychologist, 42(2)* , 99-107.

Kapur, M. (2008). Productive Failure. *Cognition and Instruction, 26(3)* , 379-424.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41(2)* , 75-86.

Klahr, D., & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science, 12*, 1-48.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning. *Psychological Science, 15(10)* , 661-667.

Kuhn, D. (2005). What needs to be mastered in mastery of scientific method? *Psychological Science, 16(11)*, 873-874.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Society for Research in Child Development Monographs 60(4, Serial No. 245)* .

Linn, M. C., & Hsi, S. (2000). *Computers, Teachers, Peers: Science Learning Partners*. Mahwah, NJ: Erlbaum Associates.

*Massachusetts Department of Elementary and Secondary Education (ESE).* (2008). Retrieved 10 2008, from http://profiles.doe.mass.edu

Sao Pedro, M., Gobert, J., Heffernan, N., & Beck, J. (2009). Comparing Pedagogical Approaches for Teaching the Control of Variables Strategy. *N.A. Taatgen & H. vanRijn (Eds.), Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 1294-1299). Austin, TX: Cognitive Science Society.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49*, 31-57.

Schunn, C., Raghavan, K., & Cho, N. (2007, April 9-13). Domain-general Learning Accelerators in Middle-School Science. Chicago, IL: Presented at the annual meeting of the American Educational Research Association.

Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills; Instructional effectiveness and path independence. *Cognitive Development 23* , 488-511.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., et al. (2005). The Andes physics tutoring system: Lessons Learned. *International Journal of Artificial Intelligence and Education, 15(3)*, 1-47.

## Acknowledgments

This research was funded by the National Science Foundation (NSF-DRL#0733286; NSF-DGE# 0742503) and the U.S. Department of Education (R305A090170). Any opinions expressed are those of the authors and do not necessarily reflect those of the funding agencies.

# A critique of how learning progressions research conceptualizes sophistication and progress

Tiffany-Rose Sikorski, David Hammer, University of Maryland, College Park, MD 20742
tsikorsk@umd.edu, davidham@physics.umd.edu

**Abstract:** Researchers in science education have moved quickly to pursue "learning progressions," defined by the NRC (2007) as "descriptions of the successively more sophisticated ways of thinking about a topic" (p. 219). Given the speed of its adoption, it is not surprising there are variations in how the notion is understood, regarding how to assess sophistication as well as how to conceptualize progress. We examine learning progressions by three leading groups, to challenge assumptions that (1) ideas are "more sophisticated" insofar as they align more closely with end-state canonical knowledge, and (2) student progress can be characterized as a sequence of levels. These assumptions conflict with advances in science education research toward views of learners' knowledge and reasoning as complex, dynamic ecologies. By moving quickly to embrace learning progressions as an organizing concept for research, the community risks surrendering its own hard-won progress.

## Introduction

In the past few years science education has turned toward *learning progressions* as an organizing concept for research. Learning progressions are "descriptions of the successively more sophisticated ways of thinking about a topic" (NRC, 2007, p. 214). Already, researchers have articulated learning progressions (LPs) for many different grade levels, topics, and aspects of science. Some LPs focus narrowly, such as on celestial motion (Plummer & Slagle, 2009). Others focus broadly, such as the LP for atomic-molecular theory (Smith, Wiser, Anderson, & Krajcik, 2006). There are also LPs focusing on students' engagement in modeling (Schwarz, Reiser, Davis, Kenyon, Acher, Fortus et al., 2009) and other scientific inquiry practices. In addition to articulating these descriptions of progress, researchers have designed a variety of supporting materials for LPs, including assessment instruments, curricula, and instructional strategies.

Work on LPs has proceeded apace, as reported in a dedicated conference (Alonzo & Gotwals, 2009), a special journal issue (Hmelo-Silver & Duncan, 2009), a chapter in a National Research Council publication (NRC, 2007), and at least one literature review (Salinas, 2009). Recently, a working panel identified and discussed learning progressions from at least twenty different research projects, in part to define a common set of guidelines for future LPs (Corocoran, Mosher, & Rogat, 2009). Some researchers are even looking ahead to ways that LPs could be used to inform curricula and professional development (see for example Corrigan, Loper, Barber, Brown, & Kulikowich, 2009; Furtak, 2009).

Given the speed at which the community has adopted the notion of a learning progression, it is not surprising that there are variations in how that notion is understood. Much in particular depends on how educators understand "increasingly sophisticated," and in this there is tension over whether to focus on disciplinary knowledge or on disciplinary practices (Hodson, 1988; Hutchison & Hammer, 2009). To focus on the former, we should consider an idea more sophisticated if it comes closer to current scientific understanding. To focus on the latter, we should consider an idea more sophisticated if it was generated and defended in ways that more closely resemble scientific practices, based on the evidence and reasoning available to students. That is, in considering what constitutes increasing sophistication, what aspect of the discipline should have priority?

Much depends as well on how educators conceptualize what constitutes progress. In this there is also tension, much of it tacit, over whether to expect a linear progression from one way or level of understanding to another, or whether to expect a variety of learning paths. Most advocates of LPs explicitly endorse the latter, but in practice published LPs more typically describe a sequence of levels.

In this paper we offer a critical review of three learning progressions, in student understanding of (1) heredity (Roseman et al., 2006), (2) force and motion (Alonzo & Steedle, 2008), and (3) matter (Carraher, Smith, Wiser, Schliemann, Cayton-Hodges, 2009; Wiser, Smith, Doubler, & Asbell-Clarke, 2009). The authors in each case are prominent researchers on learning progressions, and as a set, these LPs reflect a range of approaches and perspectives with respect to assessing sophistication and conceptualizing progress. We begin with a summary of each progression with respect to its scope, development, and substance.

## Three Learning Progressions

### Roseman, Caldwell, Gogos, & Kuth (2006): Molecular Basis of Heredity

This LP aims to present the sequence of ideas that will lead to a "coherent understanding" of the two main functions of DNA: "determining the characteristics of organisms" and "passing information from one generation to the next" (Roseman et al., 2006, p. 1). The learning progression spans grades K-12, but the authors describe in more detail the sequence of ideas they propose students encounter in high school.

The authors drew on the *Benchmarks for Science Literacy*, the *Atlas of Science Literacy,* and a series of other standards documents, as well as on their own evaluation of the discipline and education research studies, to identify potential sequences of scientific ideas that could lead to an understanding of heredity. Roseman et al. (2006) propose an order of topics that deviates from the traditional textbook approach: they present "proteins before DNA" and "DNA before genes and chromosomes" (p.1).

The LP takes the form of a map and consists of two interrelated strands corresponding to the two main functions of DNA. Both strands connect to the learning goal: students' understanding that "heritable characteristics ultimately produced in the development of an organism can be observed at molecular and whole-organism levels—in structure, chemistry, or behavior" (Roseman et al., 2006, p. 6). For example, the LP has this sequence of ideas as part of the strand for learning the second function of DNA:

> Drawing mostly from benchmarks in the Heredity section of *Benchmarks'* Chapter 5: The Living Environment, the learning progression expects students in grades K-2 to learn that offspring resemble their parents (rather than other kinds of organisms). In grades 3-5 students learn that for offspring to resemble their parents, there must be a reliable way to transfer information from one generation to the next. In high school, with their prior knowledge of cells and protein molecules, students are ready to learn about the link between proteins and DNA and, hence, between DNA and traits. (Roseman et al., 2006, p. 2)

The authors provide a similar description for students' progressively more sophisticated understanding of the first function of DNA.

Roseman et al. (2006) propose to continually refine and validate their LP. First, they will reconsider "the grain size and language of the ideas" in the LP, ensuring that they align with standards, activities, and assessments (p. 3). Second, they intend to clarify the boundaries of "what specific knowledge students are and are not expected to know" (p. 3). As well, the authors will review the literature to identify common misconceptions regarding each idea. The misconceptions will inform the design of assessments and instructional activities, as well as help clarify the language of the LP itself. Finally, the authors plan to identify phenomena that will help students overcome their misconceptions and master the ideas in the LP.

### Alonzo & Steedle (2008): Force and Motion

Alonzo & Steedle (2008) are developing a learning progression to serve as an interpretive framework for assessment. This LP is designed to "diagnose" students' understanding of force and motion, particularly after an introductory unit on the topic (p. 3).

Alonzo & Steedle (2008) define LPs as "ordered descriptions of students' understanding of a given concept" (p. 1). They employ an "iterative process" of (1) hypothesizing a series of levels for student thinking, initially based on the research literature, (2) constructing and implementing related assessments to probe student thinking at each of the levels, and (3) revising the hypothesized levels based on the results of the assessments (p. 4). The goal of the revisions is to ensure that the LP captures the breadth of student thinking about force and motion, as well as to ensure that assessment items accurately diagnose students' level on the progression:

> [T]he learning progression represents a hypothesis about student thinking, rather than a description. As such, it expresses a current idea about how student understanding develops, which can—and should—be revised in response to new information about student thinking. (Alonzo, & Steedle, 2008, p. 5)

Alonzo & Steedle (2008) used the *National Science Education Standards* for eighth grade to inform their initial hypothesis, specifically to define the highest level of the LP. They also conducted a review of the literature to compile a list of "common student conceptions" about force and motion, and ordered these conceptions into levels “based upon research literature…and (occasionally) the relative difficulty of these ideas, as well as a logical consideration of proximity to the top level of the learning progression” (p. 7-8).

The current version of the LP consists of 5 levels (0-4) and describes students’ thinking at each level along four kinds of problems: force or no force, and motion or no motion. The LP also includes anticipated common errors at each level. For example:

> Level 2
> Student believes that motion implies a force in the direction of motion and that nonmotion implies no force. Conversely, student believes that force implies motion in the direction of the force.
> Force: If a force is acting upon an object, it is moving.
> 2A: The force acting on an object could be the initial force (which is carried with the object and may dissipate over time).
> No Force: If no force is acting upon an object, it is not moving.
> Motion: If an object is moving, a force is acting upon it.
> No Motion: If an object is not moving, no force is acting upon it.
> Common Errors:
> - If there is no motion, there are no forces acting.
> - When an object is moving, there is a force in the direction of its motion.
>   - 2A: This motion could be the force that put the object into motion initially.
>   - 2A: This object may come to rest because the force it carries with it has been used up. (Alonzo & Steedle, 2008, p. 16-17)

In addition to creating descriptions of each level in the LP, Alonzo & Steedle have developed associated assessment items for diagnosing students' levels on the progression.

## The Inquiry Project: Matter

The Inquiry Project is currently conducting research on 3rd-5th grade students' reasoning about matter, material, weight, volume, and density, in part to contribute to a learning progression on the nature of matter (Smith, Wiser, Anderson, & Krajcik, 2006). The group has yet to publish a complete LP for these grades. However, two recent Inquiry Project publications (Carraher, Smith, Wiser, Schliemann, Cayton-Hodges, 2009; Wiser, Smith, Doubler, & Asbell-Clarke, 2009) describe in detail the project's work on developing such a progression.

Wiser et al. (2009) frame the problem of designing a learning progression as follows:

> [G]iven a single starting point (preschoolers’ concepts and beliefs in the matter domain) and a single target point (the atomic-molecular theory taught to adolescents in a majority of countries) in how many ways can the knowledge network evolve (when characterized at the level of concepts and beliefs)? (p. 4).

In alignment with previous work (Smith et al., 2006), the Inquiry Project grounds its work on LPs in a theoretical framework of conceptual change. In this view, students’ initial understandings of matter are fundamentally different—and incompatible—with scientists’ understandings. For example, learning the atomic-molecular theory of matter (ATM) requires massive "reconceptualization" of students' ideas about material and matter (Wiser et al., p. 2). Wiser et al. (2009) predict that there are only a few possible pathways to a learning goal such as ATM, because “knowledge network[s] can only change productively in very few ways” (p. 4). Carraher et al. (2009) anticipate that 3rd-5th grade students' progression in reasoning about matter will involve two interrelated changes:

> - A gradual shift from (a) perception-centered thinking, that is, understanding and explanation closely tied to perceptual judgment and appearances, to (b) model-mediated thinking, informed by views about matter and drawing upon a set of increasingly advanced, inter-related concepts and scientific habits of mind.
> - The development of quantitative reasoning and understanding of measurement that students can use to make predictions, interpret, and explain relationships among physical quantities... (p. 3)

The Inquiry Project has undertaken a three-year longitudinal study to explore these changes. The study involves the design and implementation of a novel curriculum aimed at "fostering the development of students' knowledge about matter" (Wiser et al., 2009, p. 1). The Inquiry Project also conducts a series of clinical interviews to explore control students' (those exposed to a traditional curriculum) and treatment students' (those exposed to the Inquiry Project curriculum) ideas about matter-related concepts (see Carraher et al., 2009).

The Inquiry Project is currently in the process of gathering and analyzing data. They have confirmed some aspects of the lower anchor of their LP, for example, that "the distinction between material kind and weight are sporadic and context-dependent" (Wiser et al., 2009, p. 9). The authors are working to establish the "intermediate" levels—or "stepping stones"—of the progression (p. 1).

## Sophistication and Progress

Despite the variety of definitions of *learning progression* circulating in the literature, they all share a general sense that LPs describe how students become more sophisticated with respect to some aspect of science (Salinas, 2009). Roseman et al. (2006) design sequences of concepts in the canon based on disciplinary knowledge as mapped by Project 2061. Alonzo & Steedle (2008) hypothesize, test, and refine sequences of student conceptions using both disciplinary knowledge and research on learning. Wiser et al. (2009) are looking for sequences of student conceptions interwoven with epistemologies and practices of science. Our focus in this critical review is on how each of these LPs assesses "sophistication" and on how each conceptualizes the dynamics of progress.

### Assessing Sophistication

It may be educational common sense that what students learn should be correct knowledge, beginning with simpler, more foundational ideas and building from there to more complex, difficult material. Children learn to count, then to add, later to multiply, and so on. Roseman et al. (2006)'s work is designed around that view of knowledge and progress. In their LP, the canonical ideas themselves constitute the pathways of learning, and these pathways are therefore naturally towards more complete, correct understandings (Furtak, 2009).

Alonzo & Steedle (2008), on the other hand, create a learning progression not from canonical ideas, but from students' ideas, both correct and incorrect. In this respect, the two LPs are quite different. Roseman et al. (2006) do not include students' alternative conceptions in their LP, a difference that we explore further in the section "Conceptualizing the Dynamics of Progress." With respect to their determinations of sophistication, however, the two LPs are similar. According to Alonzo & Steedle (2008), the force and motion LP describes

> the thinking that students at that level could be expected to exhibit, including both the correct ideas that can be carried to the next level and the misconceptions that will need to be revised before students can reach the next level (p. 4).

Like Roseman et al., Alonzo & Steedle take "more sophisticated" to mean "more correct," that is more completely aligned with the target canonical understanding.

There are several reasons, however, to question the educational common sense that student learning should proceed as a sequence of conceptual attainments. First, such a sequence has not generally characterized progress within the sciences. Unlike basic ideas of arithmetic, which have been stable for millennia, basic ideas within science have gone through dramatic change. Concepts of life, matter, and energy that are foundational today were relatively recent constructions. (That life arises only from other life and the concept of energy are less than 200 years old; the idea that matter is made up of atoms is a notable exception, although the modern understanding of atoms as opposed to molecules is quite young.)

Moreover, it has happened often in science that the formation of a wrong idea has been generative for later progress. For example, the Caloric Theory treated heat as an invisible substance that is contained in hot objects and can flow into from them to cold objects. That incorrect idea was a stepping stone toward differentiating heat as an extensive quantity from temperature as an intensive quantity, as well as toward understanding energy as conserved (Chang, 2004). That is, the formation of a wrong idea may be a productive development in science. By including only correct concepts in their LP, Roseman et al. (2006) systematically omit the possibility of productive, incorrect ideas. And although their conceptualization of LPs does not necessitate it, Alonzo and Steedle (2008) do something similar by including among their hypotheses only sequences that become more correct over time.

Second, research on science learning has come to understand that progress is not entirely, or even primarily, about correct *concepts*. Indeed, the original emphasis of misconceptions research was that students' wrong ideas are often rational in ways that are constructive of scientific practices (Strike & Posner, 1985); in subsequent years Strike and Posner (1994) specifically revised their discussion of misconceptions to emphasize that "if conceptual change theory suggests anything about instruction, it is that the handles to effective instruction are to be found in persistent attention to the argument and in less attention to right answers" (p. 171). Minds, they argued, are complex, dynamic *ecologies*. To determine the sophistication of student ideas by their alignment with the canon is to overlook other aspects of those ecologies.

One outcome of a focus on “right answers” is that, in the practices of assessment and instruction it encourages, students may learn to assess ideas by their alignment with the canon rather than by fit with the evidence and reasoning they have available (Coffey, in preparation). That is to say, measures of sophistication organized around alignment with the canon are often at odds with assessments of quality by the *practices* of science. How, for example, should a teacher think about a student's progress who develops an account of inheritance based on evidence of how a mother's behavior and health during her pregnancy affects offspring characteristics? That account may differ greatly from canonical understanding, but the student's reasoning in producing it may reflect nascent scientific practices of generating ideas from evidence and reasoning.

For a number of years, the field has moved toward conceptualizing students' progress in science in ways that consider engagement in disciplinary practices (Engle & Conant, 2002; Ford, 2005). Progress in that engagement may involve students’ pursuit of non-canonical, but by other criteria sophisticated, accounts of phenomena (Hammer, 1997; Russ, Coffey, Hammer, & Hutchison, 2009).

Unlike Roseman et al. (2006) and Alonzo & Steedle (2008), Wiser et al. (2009) work explicitly from a cognitive theory in which movement along a learning progression is more than just the linear acquisition of “more elements of the expert theory" (p. 9). Instead, they define movement along the progression such that students are put “in a better position, eventually, to understand a basic version of AMT” (p. 9). In this framework, "stepping stones" need not align with the logical structure of the end-state canonical knowledge, nor are they limited to concepts:

> [T]hey are sets of concepts, beliefs, principles, models, numerical & mathematical understandings, and representational tools that provide students with coherent interpretations of a broad range of phenomena… (Wiser et al., 2009, p. 9)

The Inquiry Group acknowledges flexibility in determining stepping stones, especially for young students. For example, they considered two potential stepping stones in the fifth grade curriculum: (i) "a solid understanding, at the macroscopic level, of weight, volume, material, mass, and density, and their interrelations, consistent with the expert view", or (ii) a particulate model of matter (Wiser et al., 2009, p. 10). The researchers opted for option (ii). That is, students are permitted to continue with their non-canonical ideas about mass and weight, for example, in order to pursue melting, freezing, evaporating, and a host of other phenomena explainable by particulate models of matter. The authors predict that pursuing a particulate model for matter will be more interesting to students, align more with students’ expectations about science, and have “the most pay-off or ‘legs’ from a scientific perspective” by “introduc[ing] students to a productive new framework for thinking about matter” (p. 11).

It is unclear, however, whether the Inquiry Project will allow students' *non-canonical models* of matter to act as stepping stones in their progression. Allowing the development of non-canonical, but productive models is not prevalent in learning progressions literature, though Stevens, Shin, & Krajcik (2009) and Corcoran et al. (2009) for example suggest that non-canonical models might help students’ move towards upper anchor understanding. Wiser et al. (2009) appear to be struggling with this aspect of the development of their LP:

> A central issue is exactly what set of elements to include in a particulate model and how to introduce them in a way that helps students understand deeper epistemological issues about models, including their tentative revisable nature and their use as tools of inquiry. (p. 11)

This is an essential challenge for research on learning progressions: Conceptual understanding is only part of a complex dynamic, and LPs that treat it as a separable component may direct educational attention in such a way as to interfere with a healthy, productive cognitive and metacognitive ecology. Understanding the mind as a complex, dynamic ecology challenges another feature of LPs as well, namely the idea that students' conceptual knowledge can be characterized by levels. We turn to that aspect of our critique now.

## Conceptualizing the Dynamics of Progress

A working panel on learning progressions recently suggested that any 'good learning progression' should contain a levels-like description of students' knowledge:

> Levels of achievement that are intermediate steps in the developmental pathway(s) traced by the learning progression. These levels may reflect levels of integration or common stages that characterize the development of student thinking. (Corcoran, Mosher, & Rogat, 2009, p. 38)

Many groups have taken levels-based approaches to constructing LPs (see for example Mohan, Chen, & Anderson, 2009; Plummer & Slagle, 2009). A levels-based LP provides an organizing structure for “grouping similar sets of ideas” about a concept together and maps a clear pathway from initial to expert understandings (Alonzo & Steedle, 2008, p. 5). For example, Alonzo & Steedle's LP consists of five levels, each describing a qualitatively distinct way that students might think about force and motion. The progression also suggests a possible pathway for "how student understanding develops"—that is, from one qualitatively distinct level to the next (Alonzo & Steedle, 2008, p. 30).

There are several reasons, however, to question a levels-based approach to characterizing students' progress in science. In their work on validating the force and motion learning progress, Alonzo & Steedle (2008) highlight one potential problem: levels may not adequately describe the state of students' knowledge.

Levels, as described by Corcoran et al. (2009), denote periods of consistency in students’ knowledge. In this view, a student who gives a Level 2 response on a Newton’s third law question, for example, should give similar responses on all Newton’s third law questions. Alonzo & Steedle (2008), however, found that students “do not respond consistently across problem contexts” (p. 29). That is, students can appear to be on two different levels simultaneously. Alonzo & Steedle attribute some of the inconsistency to ambiguities in the language of assessment items. However, the authors also acknowledge that students' reasoning may be context sensitive, and so it may not be possible to "produce a single, reliable diagnosis of a student’s level” on a learning progression (p. 29).

Steedle & Shavelson (2009) investigated whether students’ responses on the diagnostic test designed by Alonzo & Steedle (2008) do indeed “reflect the systematic application of a coherent set of ideas needed to afford valid interpretations of learning progression level diagnoses” (p. 15). They found that in general, the answer is no—students do not respond in ways that suggest systematic application of a coherent set of ideas. In other words, students’ ideas are not accurately described by coherent, qualitatively different levels. The two exceptions to this finding are students “whose understanding is (nearly) scientifically accurate and those who believe that velocity is linearly related to force” (p. 15). The phenomenology of levels-like response patterns cannot be denied in these narrow exceptions. However, both Alonzo & Steedle and Steedle & Shavelson suggest that the levels-based approach is of limited use in instances where students show evidence of “unstable, context-dependent” reasoning (Steedle & Shavelson, 2009, p. 15).

Research on learning in science has been moving toward complex, dynamic views of cognition (Strike & Posner, 1992; Thelen & Smith, 1994; diSessa, 1993; Redish, 2004; Brown & Hammer, 2008). Rather than describe students as “having” or “not having” a particular level of knowledge, this research conceptualizes students’ knowledge as manifold, context-sensitive, and coupled to and embedded in the social and physical environment. Scherr (2008) for example documented how a college student could easily distinguish mass and density in one context, but conflated them in another. Frank (2009) documented multiple stabilities in how groups of students reason about motion. Smith (2005) found that she could alter infants’ performance on the Piagetian A-not-B task by changing their body positions or adding weights to their wrists. These and other results challenge the view that students possess a static “level” of knowledge, skill, or understanding.

Evidence that student knowledge is generally not well characterized as level-like at any point in time, clearly, raises questions regarding LPs composed of a succession of qualitatively different levels of knowledge or understanding. According to Corcoran, Mosher, & Rogat (2009), the levels approach

> stems from a structural view of cognitive development which suggests that the development of student thinking may not be purely incremental but may proceed as a series of increasingly complex schemes for organizing understanding of the world which may be rather stable for periods of time, but which eventually are modified or even broken down and rebuilt to take account of new evidence and new perceptions... (p. 18)

That structural view is at odds with evidence of contextual sensitivity in student reasoning. Identifying levels of conceptual understanding becomes even more problematic when we consider views of how different aspects of learners’ cognition interact (Perkins & Simmons, 1988). Consider the challenge of describing students' progress along two just dimensions: i) content knowledge of biodiversity and ii) generating evidence-based explanations. Songer, Kelcey, & Gotwals (2009) present one solution: create an LP for concepts in biodiversity, create a separate LP for generating explanations, and then measure students’ location on each. This approach, we suggest, misses the complexity of the interaction. Lehrer & Schauble (2009) make this point succinctly in their commentary:

> We hope that Songer et al. will elaborate the meaning of complexity of explanation...We suspect that complexity interacts with the nature of the knowledge of biodiversity being assessed, and a syntactic definition may miss this interaction. (p. 732)

For an example from physics, understanding the Newtonian concept of *force* (or *mass*, or *energy*) is in close interaction with understanding the disciplinary practices of positing universal "laws" of nature, of expecting and working to construct underlying principles that govern *all* experience within a domain. Learning the concept *involves* participating in that practice; and for many students coming to a first appreciation of that practice happens in learning Newtonian mechanics. The inquiry practices and the conceptual understanding are inseparable.

## Concluding Remarks

Our purpose in this presentation has been to lay on critical brakes to the adoption of learning progressions as the organizing concept and language for research on learning. By reviewing LPs constructed by several prominent groups, we hope to have illustrated how at least some LPs assume progress occurs as a sequence of conceptual attainments, monotonically increasing toward the end-state disciplinary view. Research in science education has slowly made progress toward views of learners' knowledge and reasoning in a complex conceptual, metaconceptual, motivational and social ecology. The field should take care that the quick adoption of "learning progressions" as a construct for organizing work does not set that progress aside.

## References

Alonzo, A., & Gotwals, A.W. (Eds.). (2009, June). *Proceedings of the Learning Progressions in Science (LeaPS) Conference*, Iowa City, IA. Available from http://www.education.uiowa.edu/projects/leaps/proceedings/

Alonzo, A., & Steedle, J. T. (2008). Developing and Assessing a Force and Motion Learning Progression. *Science Education*, *93*(3), 389-421.

Brown, D. E., & Hammer, D. (2008). Conceptual change in physics. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 127-154). New York: Routledge (Taylor & Francis Group).

Carraher, D., Smith, C.L., Wiser, M., Schliemann, A., & Cayton-Hodges, G. (2009, June). *Assessing students' evolving understandings about matter*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA. Available from http://www.education.uiowa.edu/projects/leaps/proceedings/

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. New York: OUP.

Corcoran, T., Mosher, F.A., & Rogat, A. (2009). *Learning Progressions in Science. A Evidence Based Approach to Reform*. Research Report # RR-63. Center on Continuous Instructional Improvement. Teachers College–Columbia University.

Corrigan, S., Loper, S., Barber, J., Brown, N., & Kulikowich, J. (2009, June). *The juncture of supply and demand for information: How and when can learning progressions meet the information demands of curriculum developers?* Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA. Available from http://www.education.uiowa.edu/projects/leaps/proceedings/

diSessa, A. (1993). Toward an epistemology of physics. *Cognition & Instruction*, 10(2/3), 105-225.

Engle, R.A., & Conant, F.R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition & Instruction, 20*(4), 399-483.

Ford, M.J. (2005). The game, the pieces, and the players: Generative resources from two instructional portrayals of experimentation. *Journal of the Learning Sciences*, *14*(4), 449-487.

Furtak, E. (2009, June). *Toward learning progressions as teacher development tools*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA. Available from http://www.education.uiowa.edu/projects/leaps/proceedings/

Frank, B. (2009). The dynamics of variability in introductory students' thinking: Examples sfrom kinematics. Unpublished doctoral dissertation, University of Maryland, College Park. Available from http://www.physics.umd.edu/perg/dissertations/Frank/

Hammer, D. (1997). Discovery learning and discovery teaching. *Cognition & Instruction, 15*(4), 485-529.

Hmelo-Silver, C.E. & Duncan, R.G. (Eds.). (2009) Special Issue: Learning Progressions. *Journal of Research in Science Teaching*, *46*(6), 605-737.

Hodson, D. (1988). Toward a philosophically more valid science curriculum. *Science Education*, 72(1), 19-40.

Hutchison, P. & Hammer, D. (2009). Attending to student epistemological framing in a science classroom. *Science Education* doi:10.1002/sce.20373

Leher, R., & Schauble, L. (2009). Images of learning, images of progress. *Journal of Research in Science Teaching, 46*(6), 731-735.

Mohan, L., Chen, J., & Anderson, C.W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching, 46*(6), 675-698.

(NRC) National Research Council (2007). Learning Progressions. In Duschl, R. A.,Schweingruber, H. A., Shouse, A. W. (Eds.) *Taking Science to School: Learning and Teaching Science in Grades K-8*, (pp. 213-250) Washington, DC: The National Academies Press.

Perkins, D., & Simmons, R. (1988). Patterns of misunderstanding: An integrative model for science, math, and programming. *Review of Educational Research*, *58*(3), 303.

Plummer, J., & Slagle, C. (2009, April). *Children explaining celestial motion: Development of a learning progression*. Paper presented at the National Association for Research in Science Teaching Conference, Garden Gove, CA.

Redish, E.F. (2004). A theoretical framework for physics education research: Modeling student thinking. In *Proceedings of the International School of Physics, "Enrico Fermi" Course CLVI*, E.F. Redish & M. Vicentini (Eds.), IOS Press, Amsterdam, 2004.

Roseman J. E., Caldwell, A., Gogos, A., Kuth, L. (2006) *Mapping a Coherent Learning Progression for the Molecular Basis of Heredity*. Project 2061, American Association for the Advancement of Science. Available from http://www.project2061.org/publications/articles/papers/narst2006.pdf

Russ, R., Coffey, J., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education, 93*(5), 499-515.

Salinas, I. (2009, June). Learning progressions in science education: Two approaches for development. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA. Available from http://www.education.uiowa.edu/projects/leaps/proceedings/

Scherr, R. (2008). Gesture analysis for physics education researchers. *Physical Review Special Topics-Physics Education Research, 4*(1), 10101.

Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632-654.

Smith, L.B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review, 25*, 278-298.

Smith, C., Wiser, M., Anderson, C., and Krajcik, J. (2006) Implications of Research on Children's Learning for Standards and Assessment: A Proposed Learning Progression for Matter and Atomic-Molecular Theory. *Measurement, 14* (1&2), 1-98.

Songer, N.B., Kelcey, B., & Gotwals, A.W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching, 46*(6), 606-609.

Steedle, J., & Shavelson, R. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching*, *46*(6), 699-715.

Stevens, S., Shin, N, & Krajcik, J. (2009, June). Towards a model for the development of an empirically tested learning progression. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA. Available from http://www.education.uiowa.edu/projects/leaps/proceedings/

Strike, K., & Posner, G. (1985). A conceptual change view of learning and understanding. In L. West & A. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 211–231). New York: Academic.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.), *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice* (pp. 147-176). Albany: State University of New York Press.

Thelen, E., & Smith, L.B. (1994). *A dynamics systems approach to the development of cognition and action.* Cambridge, MA: Bradford Books/MIT Press.

Wiser, M., Smith, C.L., Doubler, S., Asbell-Clarke, J. (2009, June). Learning progressions as tools for curriculum development: Lessons from the Inquiry Project. Paper presented at the Learning Progressions for Science (LeaPS) Conference, Iowa City, IA. Available from http://www.education.uiowa.edu/projects/leaps/proceedings/

## Acknowledgments

The authors would like to thank the members of the *Learning Progressions for Scientific Inquiry* project for their suggestions and feedback on this paper. This work was supported in part by the National Science Foundation under grant # DRL 0732233, *Learning Progressions for Scientific Inquiry: A Model Implementation in the Context of Energy*. The views expressed are those of the authors and not necessarily shared by the Foundation.

# Dynamics of disciplinary understandings and practices of attending to student thinking in elementary teacher education

Janet Coffey, Ann R. Edwards, Carla Finkelstein
University of Maryland, Department of Curriculum and Instruction, College Park, MD, 20742
Email: jecoffey@umd.edu, aedwards@umd.edu, cfinkels@umd.edu

**Abstract**: This paper presents case studies of three elementary teacher candidates in a one-year master's certification program, focusing on their participation in science and mathematics methods courses. Through one primary case and two contrasting cases, we examine the interplay among a teacher candidate's orientation to each discipline (math and science), orientation towards discipline-specific teaching, and emerging practices of attending to student thinking in disciplinary domains. This study is of interest because it reveals significant differences between learning to teach science and math, suggesting that practices of attending have disciplinary grounding and therefore do not necessarily transfer across disciplinary contexts. We argue that better understanding these dynamics becomes important for teacher education, particularly for elementary teachers, who are responsible for teaching across disciplines.

## Background & Motivation

High quality teacher education programs emphasize student-centered curriculum and instruction (Darling-Hammond, Hammerness, et al, 2007). The importance of attending to the substance of student thinking is grounded in research on how people learn and construct understanding (National Research Council, 1999) and in the role practices of attending can play in shaping teachers' instructional moves and supporting students' learning (Ball, 1993; Black & Wiliam, 1998; Hammer, 1997). Little research addresses the role teacher education can play in facilitating the development of teachers' practices of attending to student thinking.

A recurring theme in elementary teacher education research literature concerns teachers' subject matter background and preparation; in math and science, specifically, research shows that prospective elementary school teachers often lack a strong disciplinary foundation (Ball, 1990; Lewis, Parsad, et al, 1999; National Research Council, 2007). While we know that the subject matters for instruction (Stodolsky, 1988), and that content knowledge plays a role in effective teaching practices (Hill, Rowan & Ball, 2005), the ways in which a subject matters for learning to teach is less clear. Since student thinking occurs in discipline-based contexts, better understanding how practices of attending to student thinking across the disciplines develop is a pressing issue for elementary teacher education. To that end, this study asks: How do elementary pre-service teachers (learn to) attend to student thinking and how is that process similar or different across disciplines, specifically mathematics and science?

We situate this study in elementary teacher education because of the dynamics that arise in elementary classrooms: the same teacher is typically responsible for teaching the same group of students all of the school subject areas. This poses a unique challenge for teacher education, as elementary teacher candidates must learn about content, pedagogy, student thinking and student learning in multiple disciplines. Furthermore, we specifically chose to study elementary mathematics and science teacher preparation for two reasons. First, as pressure for improving mathematics and science achievement and calls for accelerating coverage reaches into the elementary grades, high quality elementary mathematics and science teaching is critical. Second, elementary teacher candidates often have had difficult experiences with mathematics and science and come to teacher preparation with negative orientations to the two disciplines and anxiety about teaching them (Sowder, 2007). Therefore, coming to understand the ways in which their discipline-relevant experiences and backgrounds interplay with learning to be responsive to student thinking are key considerations in effective elementary teacher preparation (Kilpatrick, Swafford, & Findell, 2001).

## Our Study

Our study took place during the 2008-09 school year in the context of a one-year elementary master's certification program at a large public university located in the mid-Atlantic United States. The study specifically focuses on the math and science methods courses and associated field experiences. Two of the authors were course instructors.

Both courses framed teaching as responsiveness to students' ideas and reasoning. With this emphasis, the courses foregrounded everyday assessment as a driver for instructional decision-making (Atkin & Coffey, 2003; Erickson, 2007; Levin, Hammer & Coffey, 2009). Coursework reflected this orientation: Both courses asked students to engage in case studies of student thinking, and individual and collaborative reflection on student work, among other field-based assignments. Both courses regularly engaged students in disciplinary reasoning and problem solving; in other words, the pre-service teachers did math and science, and reflected on what that doing entailed. Each course met weekly for three hours during the Fall semester. In addition to university coursework, the pre-service teachers spent three days each week in field placements in large public school districts with uniform curricular demands and diverse student populations.

## Subjects

Our analysis draws on data from three of the 25 consenting pre-service teachers in the cohort. All three were interning in 3rd grade classrooms. Brief profiles of the case study teachers are presented here; additional information relevant to their teaching is provided below in the analysis and discussion. Together the cases capture the variation along the axes of disciplinary orientation and attending practices explored in this paper. Kim, who serves as the primary case for this paper, is a white female in her mid-20s, with an undergraduate degree in Communications. Natalia is a white, female in her mid-30s who received dual undergraduate degrees in International Studies and Russian Studies. She is a former attorney, and, just prior to the program, was a stay-at-home mother. Barbara, a white female in her late 40s, received an undergraduate degree in Psychology, with a minor in Physics.

## Data Collection & Analysis

Data is drawn from the semester-long math and science methods courses, taken concurrently, including artifacts generated in participants' field placements. Specifically, we analyzed audio and video records of course meetings as well as the course assignments, which include: written reflections and activities involving analyses of students' scientific and mathematical thinking; field-based assignments involving instructional design and implementation; an analytic case of their students' disciplinary thinking and leaning (for math and science); and observations of teaching.

Initial analysis entailed iterative coding and comparative review of pre-service candidates' work and the nature and patterns of their class participation in both methods courses. An interplay of three dimensions emerged that are particularly salient for understanding how the pre-service teachers were learning to teach in math and science:

1. **Orientation to the discipline**: Beliefs about nature of each discipline (math and science), discipline-related experiences, epistemological issues, attitudes and dispositions, content area competency (self-described and demonstrated).
2. **Orientation towards discipline-specific teaching**: Visions of discipline-specific teaching, role of the teacher, ideas about student learning, goals and priorities within the discipline.
3. **Practices of attending to student thinking**: The aspects of student thinking that participants notice, what sense they make of that, how they take that up in their analyses and recommendations for next steps for teaching and learning.

Our analysis in this paper reflects case study methods used to capture variation in dynamics among these dimensions. We selected three cases that illuminate important distinctions across individuals as they learned to teach science and math and reflect patterns of variation that we observed among the broader group of 25. Kim serves as a primary case study to organize data and discussion. The other two cases presented here–Barbara and Natalia–offer interesting contrasts, which we discuss at the end of this paper. We are not arguing that these are the only dynamics at play; however, across the pre-service teachers, these dimensions emerged as fundamental to understanding the variation in learning to teach in discipline domains. Of course, school context played a role in shaping teachers' expectations and what they did in the classroom. While all three taught the same grade in large public school systems in the same state (thus sharing state standards), Kim and Natalia taught in a district that, overall, was higher performing than the district where Barbara taught.

# A case study of elementary pre-service teachers' engagement in attending to student thinking in math and science – a closer look at Kim

Like many of her peers, Kim demonstrated differences in her practices of attending to student thinking across her math and science methods courses. In her science methods class, attending to student thinking afforded Kim opportunities for deep engagement in disciplinary practices and in student reasoning. In math, while she clearly demonstrated personal engagement in mathematical inquiry and problem-solving, the purposes Kim ascribed to attending to student thinking—to assess students' progress toward successful problem solving—constrained how she inquired into student thinking and what that attention made available for her. We found that Kim's differing views of the disciplines, as well as the differing ways in which the presses (e.g. curricular, external testing demands) manifested in math and science teaching in her field placement setting, contributed to marked differences in her abilities to attend to students' thinking in mathematics as opposed to in science. Before we take a closer look at the ways in which she attended to student thinking, we briefly describe her orientations to the disciplines of math and science and towards teaching in these disciplines.

## Orientation to the discipline

In science methods, Kim initially portrayed science as wonderment and exploration on one hand and "knowledge" on the other, without a clear articulation of the relationship. Her work over the semester in attending to students triggered her own episodes of engagement in scientific reasoning. Kim began to appreciate scientific reasoning as a way to get from wonderment and exploration to knowledge. Scientific reasoning took on dimensions of instrumentality—she began to see it as a useful "tool" for asking questions about the world.

During the semester in her math methods class, Kim's orientation to mathematics shifted from a belief that mathematics is primarily about getting a "right answer" and a focus on its instrumentality to the centrality of conceptual understanding and sense-making. At the start of the semester, Kim reflected that her experiences learning mathematics fostered a view of school mathematics as concrete, straightforward exercises that "are either right or wrong" and a view of mathematics learning as primarily mastery through repetition and memorization. She recognized the instrumentality of mathematics outside of school and felt that the usefulness of mathematics was a key motivation for students, including herself, to learn mathematics. Through the course of the semester, doing math became about achieving deep conceptual understanding and the satisfaction of "getting it" through "successful" problem solving. She approached mathematical explorations and problematic tasks in the methods class as puzzles to be pursued. Despite her clear enthusiasm and competency in pursuing these personal mathematical challenges, her central goal, however, remained "getting" or mastering them as opposed to the processes of conjecturing or inquiry.

In both disciplinary areas, Kim demonstrated a curiosity toward puzzling situations and problems (including self-initiated questions) that support deep engagement in disciplinary practices. However, in math her primary focus was on understanding achieved through successful problem-solving, whereas in science she demonstrated a value on the processes of questioning and reasoning unto themselves.

## Orientation to discipline-specific teaching

During the science methods course, Kim's focus shifted from student ownership through topic relevance to creating opportunities for students to articulate and question their scientific reasoning. Her primary goal involved helping students engage in questioning and sustained reasoning of everyday phenomena to which they may or may not bring prior knowledge, or even see as relevant. To help students engage with each other's ideas, Kim regularly asked, "Do you understand what he is saying?" and "Does that make sense?" Science content goals became secondary in practice to her facilitation of and responsiveness to reasoning, although she did not drop them all together. For her, one would lead to the other.

Teaching math for Kim involved providing students with opportunities to explore and make sense of rich problems in order to achieve "success" on the problem and, thus, develop understandings of mathematical concepts. Her role was to reveal students' strategies and reasoning by asking questions and then use that information to pose guiding questions to support students' movement toward successful solutions. During the math methods course, she also began to value students' sharing of their strategies with one another to make multiple solutions available to students.

## Practices of attending to disciplinary thinking: Kim's science teaching

Over the course of the semester in her science methods coursework, the focus of Kim's analysis of classroom activity was on the substance (beyond canonical correctness) of students' ideas and reasoning. In her classroom teaching experiences, attending to student thinking served as a way for Kim to help her students clarify their reasoning and engage with each other's ideas. Kim's written analyses show that she began to see her role as a science teacher as one of facilitating students' articulation of reasoning as a bridge to scientific understanding. The analysis below helps illuminate this positioning.

The following excerpt comes from Kim's transcript of her 3rd grade class discussing why days were longer in summer than in winter. This segment begins with an explanation set forth by one of the students, that "the sun goes around the Earth slower and the Earth spins faster." On first consideration, what may jump out in the student's statement are the inaccuracies of the students' ideas. Kim's treatment of those ideas–both in her subsequent moves in the classroom and her written reflection—demonstrates a commitment to helping students clarify their reasoning and engage with each other's ideas. Despite the fact that Kurt's and Rye's ideas were wrong, Kim sought clarification, in a way that indicates she was trying to understand the reasoning behind the explanations.

Kurt: I was thinking that the sun…that the sun goes around the Earth slower and the Earth spins faster. So, so, it's um, so the sun goes to the other side of the Earth, um it goes to the other side of the Earth cause because the Earth is rotating around…(trying to show what he is saying with his hands)

Kim: Can you draw me a picture of what you mean up here? Can you draw a picture? I'll erase some space. Can you draw a picture you think?
Kurt: Yup.
Kim: Ok draw us a picture. Let's see if we agree on understanding Kurt.
Kurt: So this (*drawing suns orbit around Earth*), so the sun is going around and this the Earth is spinning fast. So this is the, hold on, this is the Northern Hemisphere and this is the Southern. So now it's so, so, and then it turns like it turns really fast like, and if you, its night time and the sun is here and…
Kim: Ok. Guys, look up here at what Kurt did. He said, here's the sun, he said the sun is going around the Earth like this. And the Earth is spinning like that. Does anyone agree or disagree with what Kurt is saying. Think about what we know about the Earth, and day and night, and the sun…Rye?
Rye: I agree.
Kim: You agree with him? Why do you agree with him?
Rye: Because the sun does go around fast, and the Earth, but I disagree a little bit because the Earth goes slow around.
Kim: Explain it, say it again. What do you mean the Earth goes faster and what do you mean it goes slower?
Rye: It goes slower because if it went faster it would be the year 3000 tomorrow.
Kim: So you're saying if it went too fast then time would pass?

In her analysis of this conversation Kim wrote, "Some students said that the Earth can speed up and slow down, and that's why we have longer or shorter days…They do get that the Earth rotates on its axis and that is what causes day and night. So it must seem natural that the speed of the rotation must change to explain the change in the length of day. We can't feel the Earth moving anyway, so we probably wouldn't feel a change in speed…Looking back I could have followed up with asking them what causes the Earth to spin slower or faster." We see her pushing students to consider the ideas in light of other things they "know" – "think about what we already know about the Earth, and day, and night, and the sun." Her analysis indicated that she was also able to identify instances where she could have pressed students on their reasoning, and in particular push for a causal explanation.

Kim appeared to allow the specifics of her conceptual understanding goals to emerge from the discussion. For example, although she is expressed concern about the inaccuracy of the statement that "the sun goes around the Earth slower and the Earth spins faster," Kim addressed students' conceptual understandings not by "correcting" their statements but by encouraging students to examine the logic of their reasoning. In any given moment of interaction, she primarily attended to students' clarity of articulation and reasoning, allowing students to engage meaningfully with each other ideas.

Following the exchange in the transcript above, Kim offered students a globe and differently sized balls "to allow them to model what they knew." She explained, "I had hoped that if we established some things we all agreed to be true, we would be able to reason that the Earth always rotates on its axis at a constant speed once every twenty-four hours." A student then challenged the idea that the sun orbits the Earth, offering as explanation that "the Earth spins so slow that the side facing the sun is in the summer and it stays that way until it moves out of the light, and then it becomes winter." Kim noted, "Her response was right in so many ways…she knew that the Earth orbits around the sun and that the relationship somehow caused the different seasons to happen on different sides of the earth. She also knew that the Earth was rotating on its axis while it was orbiting. However, it was hard for me to let it go there because we just established that the day/night cycle happens every twenty-four hours…"

Kim held students accountable for their scientific reasoning. Recognizing the inconsistency in the student's reasoning, Kim explained: "I brought the discussion back to that idea and sure enough a student was able to challenge her on her idea. Dan responded, 'I have to disagree with Verna because…if it was summer over here, if it was summer over here, then like how could it move because its summer over here and its winter over here. I don't get it. Because if it was summer over here, it would have to be summer there for, like, a season and if was, um, winter over here it would have to be winter for a season. And then they would never have day, well it would never be night.'" As this brief snippet exemplifies, over the semester Kim began to see reasoning as the bridge to scientific knowledge, and her role as teacher as one of facilitating this connection. For Kim, the end point of any particular science lesson was less important than the process by which students got there, which she came to see as the primary work for learners doing science. While this was the case within a particular lesson, at the end of a several lessons, Kim expected to see conceptual progress towards canonically accepted explanations. Yet, she was able to let this end goal go in most moments of interaction.

## Practices of attending to disciplinary thinking: Kim's mathematics teaching

Kim's attending practices looked different in the math methods course. Her analyses of students' mathematical thinking addressed conceptual and procedural understandings as she attended to the details of students' talk, inscription and gesture to make her judgments. She increasingly saw the value of revealing student thinking, but viewed attention to student thinking in math as a teaching strategy for assessment of student understanding and readiness. In her practice, Kim questioned her students primarily to determine their problem-solving strategies and, if necessary, to steer them toward a successful solution (while honoring and building upon the particular direction each student had attempted to take).

The following example comes from Kim's instruction with a small group on the task: "Greg climbed 2,600 meters up the side of a mountain. His brother Harry climbed half as far. How far did Harry climb?" Before starting the task, Kim focused the students on problem-solving strategies ("What are some important steps when you are problem solving?" "What are some possibilities for solving problems that maybe you've tried before, some strategies you used?") as well as the multiplicity of strategies that could be used to solve a problem ("Are there different ways to solve problems?" "Do you agree that there are multiple ways to solve a problem not just one correct way?"). Her interactions with the students on this task reflected how she uses questions to reveal students' understandings and then "guide" students to "success" on the problem. The following excerpt is typical of the kind of revealing and guiding questions she asks as well as her emphasis on "making sense."

Kim: What if we separate these two numbers [2000 and 600], Mish. What if we separate them by what their place value is?
Mish: What do you mean separate them?
Kim: What does the two really mean there?
Mish: Two thousand.
Kim: Ok, so if we do 2000, then what's the rest? Do you remember when we did expanded notation?
Mish: Yes…
Kim: So, how can we divide all these numbers in half? What is half of zero?
Mish: Zero.
Kim: Yup, half of zero is zero. What is half of 600?
Mish: 300.
Kim: You're right. How do you know that it's half of 600?
Mish: Because 300 plus 300 is 600…[Mish completes the problem.]
Kim: You're right, now tell me why.
Mish: Because that's in the thousands, that's in the hundreds, that's in the tens and ones, so and I don't have anything to add up to these so they stay in, these two I don't change.
Kim: And you're going to add these two numbers together?
Mish: Yes
Kim: So we've figured out half and you're going to add the halves together. Does that make sense to you?
Mish: No.
Kim: No? Well why doesn't it make sense. Think about it. You know you're right and I just want you to think about why that's true.

For Kim, teaching math involved questioning strategies that guide the students down solution paths: "By [this] task, I felt really comfortable with my questioning strategy and I think I was able to do less work and let the students do more work. Some of my favorite questions were: Where are you going next? How will you know when you have your answer? What does that mean? When I was watching video, I think that these questions turned the work back on the students and made them think about what it was that they were doing and trying to accomplish. It also helped me understand them."

Kim's analyses of her students' mathematical thinking address conceptual and procedural understanding, though not mathematical processes such as explanation and justification. In her analysis of the excerpt above, she wrote: "He was unable to come up with the correct answer, but the process he was using told me he understands a few things; what half means, how division and subtraction work together, and place value. He knew that half meant dividing something into two equal parts. He told me right away that half of 600 is 300. He also was using the division symbol and attempting to divide the large number by two. Later when I prompted him to check his answer he did so by subtraction and subtracted his answer from the whole number. When the answer was not equal after subtraction he knew it was the wrong answer…I know he understands place value because after attempting to divide 3,600 by 2, he broke the number down into 3,000 and 600. He still got confused because he kept putting the number back together before he knew he had found half of each."

Kim viewed attending to student thinking in math as a teaching strategy for assessment of student readiness: "I have learned that this [analyzing student thinking] is all a very important process in teaching. I have tried to incorporate this into each math class. I have found that I am able to assess student progress easier. I can tell what they might need more work on and when they are ready to move on. It also helps in giving me ideas for lessons."

In summary, in math, Kim focuses on students' problem-solving as a means of providing her information about progress and uses questioning to "guide" them toward the correct answer. In science, she attends to the nature and clarity of reasoning and ideas, largely so students can begin doing the same. In the moment, she does not appear to be concerned about the specific directionality of the reasoning (although in retrospective reflection, she comments on progress toward conceptual knowledge).

Overall, for Kim, teaching science was about fostering sense-making and teaching math was about guiding students in making sense. In math she valued reasoning toward a particular end--the problem solution. In science the particulars of the end depended, in large part, on the emergent lines of student reasoning. The differences we saw in her math and science attending practices with students seem to coordinate with differences we see in her orientations to, and relationships with, the disciplines of science and math, and teaching in the disciplines. Kim identified science as one of her favorite subjects growing up, in part because she said it was a "break" from other academic subject matters; she wrote of it as "a chance to explore, and have fun doing it." Not unlike some of her peers, she drew on experiences outside of the classroom to inform her views about science. When asked to describe a positive learning experience in science, she wrote of exploring the woods and streams around her house, and about excursions with her father. The theme of exploration – of things and ideas – is echoed in her own teaching. In mathematics, Kim's orientation shifted from a belief that mathematical activity entails getting right or wrong answers and a focus on its instrumentality, and she began to recognize and embrace the sense-making nature of the discipline. This was reflected in her practice and in her practices of attending, which focused on fostering sense-making with her students while still retaining a strong focus on successful problem-solving.

In both disciplines, Kim's practices of attending to student thinking further facilitated her own conceptual knowledge and exploration. For example, in the science methods course, discussion of video clips of elementary students doing science presented opportunities for engagement in disciplinary reasoning practices. An example occurred during a conversation the pre-service teacher candidates were having about a transcript of 2nd graders were talking about whether a cup full of water and ice will overflow when the ice melts. Two competing ideas were under consideration by the elementary-aged students: "Ice gets smaller when it melts so it would be less water;" and "So the water holds the ice up and then when it melts it sinks to the bottom and pushes the water out." (Restated as, "It sinks. It melts. And makes more water…and pushes the rest out."). Kim asked her classmates, "Why does ice expand when it freezes? What's happening?" She made several moves similar in nature over the semester. In both methods courses, she took up opportunities to analyze student thinking as invitations to engage deeply in the conceptual content. As she attends to and explores student thinking, she is making sense of student ideas *and* the phenomena about which they are reasoning, as though they are part and parcel of the same thing.

## Learning From Other Cases – Barbara and Natalia

Kim's trajectories in learning to teach math and learning to teach science were different than some of her peers. Broadly speaking, Barbara's orientation to teaching math was a strong response to her own negative experiences as a student in "traditional" math. She articulated that math teaching should be engaging and hands-on in order to motivate students and support understanding. Barbara also articulated the importance of students learning that there is more than one way to solve a problem. In her own mathematical activity, she was willing to engage the tasks and discussions about the mathematics, but displayed a lack of confidence, deferring to others and joking about her competence. Her analyses of students' mathematical thinking attended to both conceptual and procedural understandings in a detailed manner, and her lessons provided opportunities for students to cooperatively engage in tasks. However, in practice, she focused on getting students to a "right answer" through leading questioning (which she described as "guided invention"); also, her teaching reflection revealed that she was primarily concerned with students being "on-task" and enthusiastic and their arriving at correct answers.

On the other hand, Barbara reported to "love science" and to want to "inspire" that same sentiment in her students. Barbara minored in physics as an undergraduate and was the only student in the program who had significant undergraduate education in science or math. In her methods class, she positioned herself, and was positioned by others, as an expert in science. Peers often (physically) looked to her for explanations or to confirm the (in)correctness of an idea. Her disciplinary view of science revolved around canonical knowledge. Barbara believed background knowledge was a precursor to doing the work of science, which entailed controlled experimentation. She saw science as a difficult subject to learn because students "often lack background knowledge" and were coming from homes where "parents are not comfortable with their level of knowledge." On multiple occasions in the methods class, Barbara chose not to engage in scientific activity – in the sense making - because she didn't "know enough about people's background knowledge" and didn't want "to give away answers". In attending to student thinking in her field placement, she focused on correctness of knowledge, as articulated through use of appropriate terminology.

The case of Barbara highlights that disciplinary background – and confidence - alone is not enough to precipitate robust practices of attending to the substance of student thinking. Unlike many elementary school teachers, Barbara was neither math nor science phobic. In science, her academic background and self-proclaimed expertise influenced her (dis)engagement in scientific sense making and her focus in attending to student thinking, or more aptly stated, student knowledge; knowledge was paramount to reasoning. In contrast, in math, while comfortable, Barbara was not as confident. While she aspired to reform-based teaching and learning, her teaching looked more traditional, with a focus on correctness.

Natalia, our final case presented here, articulated that math should be taught "as inquiry not as conclusion" and that "discussion and experimentation are essential with each student exploring different strategies and sharing..." Rich tasks and appropriate manipulatives are also important for "helping students to understand and conceptualize abstract math concepts." She demonstrated strong mathematical competence and a disposition toward inquiry in her own math activity. During the course, she began to situate her understandings of math teaching and learning in her developing role as a teacher in a real school context, accountable to and for her students' learning. She was concerned with allowing students to harbor misconceptions and felt that exploration should be followed by "closure" in which concepts are made explicit through a teacher-led discussion of student-generated problem solutions. She was explicit about the need to address "efficiency" in a climate of accountability and strived to find "balance" between exploration and direct instruction.

Natalia wrote of science as "a process of discovery and developing knowledge and understanding about the world," which entailed systematic observation and exploration. At the start of the semester, she highlighted as a hurdle that "students are not used to following the scientific method to create their own answers." Initially, teacher content knowledge was important for "answering questions", fostering "intelligent debate", and identifying appropriate content goals. Substantive analyses of student thinking came easy to her. While her analysis foregrounded students' conceptual ideas, she also noticed reasoning that strung the ideas together. She viewed her responsiveness to student thinking as a necessary means to help students reason in order to appropriate conceptual knowledge and achieve conceptual understanding. She valued ambiguity and "wrong ideas" for information she could glean and as opportunities for students to reconcile competing ideas. Student reasoning, thus, served as a pedagogical role to inform her teaching, and not necessary as a vehicle for her students to do science.

This case highlights the dynamics for a candidate who entered the teacher preparation program with sophisticated notions of disciplinary teaching and strong analytic skills. Natalia saw immediate benefits of attending to student thinking to further align disciplines and disciplinary teaching. This initial position allowed her to meaningfully tackle both understanding and negotiating how these threads come together in the broader context of schooling. Over the semester, the alignment between her orientation towards the discipline and disciplinary teaching became more refined. Contrasts in science and math teaching that emerged were rooted in different contextual presses, such as high-stakes testing and degrees of prescription of the county curriculum.

## Conclusions

This work has begun to explore the role of disciplines (i.e., disciplinary understandings, beliefs about the disciplines, dispositions toward the disciplines) in learning to teach within and across the disciplines at the elementary grade levels. As the cases of Kim, Barbara and Natalia reveal, practices of attending to student thinking have disciplinary grounding and therefore should not be expected to share generic characteristics across disciplinary contexts. The role that discipline plays in learning to attend to student thinking, and thus, learning to teach, is important to understand, particularly for work with elementary school teachers. To support prospective teachers' learning to conceptualize and enact attending to student thinking as central to teaching, we need to better understand how their disciplinary understandings and orientations toward teaching specific to those disciplines interact with practices of attending to student thinking, including how, to what, and for what ends. We have begun to identify dimensions that seem salient and dynamic. Our work suggests that teacher education, particularly for elementary teacher candidates, should consider how the interactions amongst these dimensions of teacher knowledge and practice can be leveraged to foster deep reconsideration of discipline and disciplinary teaching. We see promise in better understanding these dynamics for informing teacher education.

This work also closely relates to notions of teacher identity and problematizes the uni-dimensional treatments of teacher identity, particularly with respect to elementary teachers. While, on some level, prospective teachers are constructing identities as 'teacher', as discussed in literature (Cohen, 2008), our work suggests that elementary school teachers may be navigating multiple identities as 'teacher' (Smith, 2007), specifically ones that reflect the disciplinary groundings of school subject matters. Policies geared to math and science teaching and common "wisdom" (i.e., constructs of "science and math" person or strong analytic thinker) often assume close alignment between notions of mathematics and science; however, experiences with and analyses of our cohort of pre-service teachers challenge this assumption, as the cases above demonstrate significant differences across these disciplinary contexts.

We situate this work in elementary teacher education and learning practices of attending to disciplinary thinking. A criticism of elementary school teaching is that teachers often lack disciplinary expertise (Ball, 2000; NRC, 2007; Seaman & Szydlik, 2007), which in the case of math and science, is often judged by the number of completed math and science courses. We do not intend to argue – even implicitly – for subject matter specialists at the elementary grade levels. While we agree that teachers often come into teaching with negative views and varied success, we question the metric of course completion as an indicator of the depth and usefulness of knowledge that can be reworked for understanding and supporting students' thinking (Ma, 1999). Moreover, we see promise in teacher education helping prospective teachers establish a stance towards practices of disciplinary attending that could contribute to more robust notions of disciplinary learning and understanding and deeper conceptual understandings – for themselves as well as for their students. We see value in teaching across multiple school matters and contexts. Knowing the student in multiple contexts, including across subject areas, could provide useful insights for identifying resources students' bring to their subject matter learning. We hope to pursue these connections in the future.

## References


Atkin, J.M. & Coffey, J.E. Eds. (2003). *Everyday Assessment in the Science Classroom*. Arlington, VA: NSTA Press.

Ball, D. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *Elementary School Journal, 93(4),* 373-397.

Ball, D. (1990). The mathematical understandings that prospective teachers bring to teacher education. *Elementary School Journal, 90,* 449-466.

Ball, D. L. (2000). Bridging practices: Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education, 51,* 241-247.

Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education, March, 1988,* 7-74.

Cohen, J. (2008). "That's not treating you as a professional": Teachers constructing complex professional identities through talk. *Teachers & Teaching, 14(2),* 79-93.

Darling-Hammond, L, Hammerness, K., Duffy, H., Lepage, P., Bransford (Eds.) 2007. *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do.* San Francisco, CA: Jossey-Bass Inc Publisher.

Hammer, D. (1997). Discovery learning and discovery teaching. *Cognition & Instruction, 15(4),* 485-529.

Hill, H., Rowan, B. & Ball, D. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal, 42(2),* 371-406.

Erickson, F. (2007). Some thoughts on "proximal" formative assessment of student learning. *Yearbook of the National Society for the Study of Education.* G. D. Fenstermacher, National Society for the Study of Education. 186-216.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Lewis, L. Parsad, B., Careym, N., Bartfai, N., Farris, E. and Smerdon, B. (1999). *Teacher Quality: A Report on the Preparation and Qualifications of Public School Teachers.* Washington DC: National Center for Education Statistics.

Levin, D.M., Hammer, D, & Coffey, J.E. (2009). Novice Teachers' Attention to Student Thinking. *Journal of Teacher Education, 60(20),* 142-154.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States.* Mahwah, N.J.: Lawrence Erlbaum Associates.

National Research Council. (2007). *Taking Science To School: Learning and Teaching Science in Grades K-8.* Washington DC: National Academics Press.

National Research Council. (1999). *How People Learn: Brain, Mind, Experience and School.* Washington DC: National Academies Press.

Seaman, C., & Szydlik. J. (2007) Mathematical sophistication among preservice elementary teachers. *Journal of Mathematics Teacher Education 10(3),* 167-182

Smith, R. (2007). Developing professional identities and knowledge: Becoming primary teachers. *Teachers & Teaching,* 13(4), 377-397.

Sowder, J. T. (2007). Mathematics teaching and classroom practice. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 157-223). Charlotte, NC: Information Age Publishing.

Stodolsky, S. (1988). The subject matters: Classroom activity in math and social studies. Chicago, IL: University of Chicago Press.

# Using Digital Video to Investigate Teachers' In-the-Moment Noticing

Bruce L. Sherin, Miriam Gamoran Sherin
Adam A. Colestock, Rosemary S. Russ
Melissa J. Luna, Martha Mulligan, Miriam Gamoran Sherin, Janet Walkoe
Northwestern University, School of Education and Social Policy
2120 Campus Drive, Evanston IL, 60208 USA
Email: bsherin@northwestern.edu, msherin@northwestern.edu, a-colestock@u.northwestern.edu, r-russ@northwestern.edu, mluna@northwestern.edu, martha-mulligan@northwestern.edu, janetwalkoe2011@u.northwestern.edu
Rogers Hall, Vanderbilt University, rogers.hall@vanderbilt.edu

**Abstract:** Understanding teacher cognition – and in particular teacher noticing - poses significant challenges to researchers because of the ongoing nature of teaching. To overcome those challenges learning science researchers have used a variety of methods to help them understand how teachers reason and make decisions while teaching. In this symposium we describe a new digital technology that allows teachers to capture short video clips of classroom activity as they are teaching. The three papers explore different aspects of what we have learned about teacher noticing using this new technology. The first paper describes the potential of the camera to serve as a window into teachers' in-the-moment noticing. The second examines teacher noticing of their students' thinking. The third paper investigates our efforts to use clips captured using this methodology as the basis for video clubs in which teachers watch and discuss excerpts of one another's classrooms.

## Introduction

Learning science researchers have, for many years, had a strong interest in teacher cognition. In particular, we would like to understand how teachers reason and make decisions in the act of teaching. Yet studying teacher thinking during instruction poses challenges for researchers. Because of the ongoing nature of teaching, it is not feasible to interrupt and ask teachers what they are thinking, what decisions they are considering, or what questions they have at the moment. Instead, researchers generally use other, less intrusive, approaches. For example, some researchers draw on observations of teaching to make inferences about a teacher's thinking (e.g., Schoenfeld, 1998). In this approach, researchers may try to construct a plausible chain of reasoning events that aligns with the teacher's actions and utterances during a lesson. Another approach involves asking teachers to reflect, after-the-fact, on their teaching. Using retrospective interviews, for instance, researchers may ask teachers to recall how they came to a particular decision or what alternatives they were considering at the time (Peterson & Clark, 1978). A third approach is to try to simulate teaching in some manner, for example, by showing teachers videos of instruction and asking them to comment (e.g., Copeland et al., 1994).

Each of these approaches, however, has limitations. Some of what a teacher thinks may not be visible in a teacher's actions or words, making it difficult to infer how a teacher is thinking simply by watching them. Also, in retrospective interviews, there is a risk that teachers will construct new interpretations of what took place during the interview, rather than recall their thinking during instruction. Similarly, even if simulations are carefully constructed to present teachers with situations that closely resemble instruction, teachers will lack the kind of detailed background information they use to make sense of real-world situations while teaching. It seems possible then that simulations will provoke a different kind of reasoning than takes place during instruction.

Whether or not these approaches suffice will depend in part on what aspect of teacher cognition the researcher is interested in. In our work, we are particularly interested in teacher noticing — in how teachers attend to and make sense of significant interactions during instruction. The classroom is a complex environment with many things happening at once. How does a teacher decide where to focus his or her attention? This issue is particularly important in light of efforts to reform mathematics and science education in the U.S. today. Teachers are called upon to continually assess the ongoing nature of lessons and to base their instruction, at least in part, on the ideas that students raise in class. In this context, how teachers direct their attention in the act of teaching is a critical area for researchers to investigate.

Given our interest in teacher noticing, the three approaches outlined above pose methodological challenges. First, it may be particularly difficult to observe a teacher "noticing;" noticing may not align with specific utterances or actions on the part of the teacher. In addition, because noticing is likely to take place in a fleeting manner it may be hard to capture retrospectively. Furthermore, during simulations noticing may operate in a manner that is quite different than how it operates during instruction. How then, can we effectively study teacher noticing?

The purpose of this symposium is to discuss a new technology that we are using to study teacher noticing during the act of instruction. Our claim is not that this new technology should replace the methods discussed above. Rather, we believe that this new technology has the potential to add an additional, complementary, window into teacher thinking in action.

### Exploring New Technology

Recently, a new class of video cameras have been introduced, cameras that are small enough to be mounted on a hat or a pair of eyeglasses. Of particular interest are cameras that offer "selective archiving" capabilities. Selective archiving allows the user to select moments of video to capture immediately after they occur.

Two cameras were used in the studies we present here: the Deja View CamWear 100 (Reich, Goldberg, & Hudek, 2004) and the POV 1.5 (V.I.O., 2009). Both cameras include two components: a small wearable camera approximately one inch long, and a recording module about the size of a cell phone that can be attached to a belt. The cameras record in a loop mode, and record over themselves after a short time. Pressing the "save" button on the recording module interrupts this process and saves the previous "loop" of video in a digital file — 30 seconds in the case of the Camwear 100 and 3 minutes in the case of the POV 1.5. In both cameras, the digital files are stored on a video card that is housed in the recording module. The files can easily be downloaded and viewed on a computer.

We developed the following two-part methodology to support our use of the new cameras. First teachers were asked to wear the camera during instruction and to press the record button "when something interesting happens." Occasionally we used variations of this prompt, but in general we asked teachers to focus on events that they found "interesting." In order to capture a complete record of the lesson taught, we also videotaped the classroom from the back of the room using a single stand-alone video camera. Second, on the same day, a researcher met with the teacher to review the set of clips that had been captured. To do so, we developed a process intended to increase the possibility of teachers drawing on their thinking during instruction, rather than creating retrospective accounts of their noticing. Specifically, a still image of the start of the captured clip was shown to the teacher. If the teacher could recall, just from the still image, why he or she captured the clip, then the video clip was not reviewed further and the teacher explained at that point why the clip had been captured. If the teacher could not recall why the clip was captured, then the clip was played but only until the point at which the teacher was able to recall his or her reason for capturing the clip. Teachers were also asked to generally describe their experiences using the camera that day, and whether the clips represent the events that the teachers had intended to capture. All interviews were videotaped and later summarized and partially transcribed.

The three papers presented in this symposium explore different aspects of what we have learned about teacher noticing thus far using this new technology. The first paper describes what we have found about the potential of the camera to serve as a window into teacher noticing during the instruction. The second paper looks closely at those clips teachers describe as being about student thinking. The third paper investigates our efforts to use clips captured using this methodology as the basis for video clubs in which teachers watch and discuss excerpts of each others' classrooms. While all three papers share methodology for studying teacher noticing, the specific data examined are somewhat different across the papers.

The format of the symposium will be as follows. We will begin with a brief introduction of the goals of our project as well as a demonstration of the wearable video camera and illustration of the kinds of data that were collected. Following this, there will be three 12-minute presentations of the papers. Next our discussant, Rogers Hall, of Vanderbilt University will provide comments on the papers. Dr. Hall's research explores the relationship between one's perception of events and how one participates in such events in a number of contexts including teaching. Finally, we will have 30 minutes for questions and discussion with the audience.

## Freezing Time: What Mathematics and Science Teachers "See" While Teaching

Bruce L. Sherin, Miriam Gamoran Sherin

Our methodology was designed to tap directly into teachers' in-the-moment noticing in a way that has not been possible before. Specifically, the idea is that the clips themselves have the potential to reveal the kinds of events that teachers pay attention to during instruction. In addition, the hope is that teachers' comments in the interviews will provide valuable information concerning the ways in which noticing acts during instruction, for example, the extent to which noticing is a conscious process and the reasons behind teachers' attention to various kinds of events. The goal of this presentation is to summarize and illustrate the data we have collected thus far, and then discuss the ways in which the methodology achieved its potential as well as problems we faced in the process.

Over the past two years, we have worked with 12 high school mathematics and science teachers who volunteered to use one of the cameras during instruction. The teachers taught in three different school districts in the Midwestern United States, all of which have diverse student populations. Most teachers used the cameras on three separate occasions; in all we conducted 39 interviews lasting approximately 45 minutes each. Each of these interviews was conducted after a teacher had worn a camera for one class session.

Overall, we found that teachers were able to use the camera to capture moments of instruction during teaching. About one-half of the teachers initially faced logistical problems using the camera — pressing the wrong button, or not pressing the button correctly, or other malfunctions of the camera. After one or two attempts at taping, however, these issues were generally resolved. The clips themselves illustrated a range of types of classroom events — whole group discussions, small group work, individual seatwork, and student presentations. Teachers reported to us that using the camera was not overly disruptive, either for themselves or for their students. This is also evident in the fact that teachers typically captured clips throughout a lesson; thus it was not the case that teachers captured clips only during the first 10 or 15 minutes of class at which point the demands of instruction took priority and camera use fell off.

On average, teachers captured 9 clips per 50 minute class period. We see this moderate number as suggesting that teachers were somewhat selective in capturing moments of instruction. Furthermore, the distribution of clips was not evenly spaced throughout a lesson. Teachers did not simply press the button every five minutes or so without regard to the specifics of the moment. Instead, teachers seemed to be sufficiently conscious of what they viewed as interesting to be able to push the button when an event stood out to them. Mason (1998) writes of the need for teachers to be "aware of their awareness." It is this kind of conscious awareness that we believe we have accessed with this methodology. Along the same lines, in many cases, teachers were able to recall what they had captured and why from looking only at the still image or a few seconds of video. Specifically, issues of student thinking, pedagogical techniques, and organizational issues were reported to have captured the teachers' attention.

Our methodology was not without problems, however. At times, teachers stated that they did not remember why a particular moment was captured. "I don't know why I pushed the button there, but I know I actively did because I can see myself [look down to press the button.]" In such cases, our methodology failed to tap into teachers' in-the-moment thinking about the noticed moment. Another concern is that there is some evidence that the act of wearing the camera and capturing moments may alter the very noticing that we intended to access. In particular, a few teachers suggested that using the camera heightened their sensitivity to noticing events that took place during instruction. "[Using the camera] made me more aware of what I thought was important." Thus while our efforts to access teachers' in-the-moment noticing appears to have been somewhat successful, there is reason to also be cautious as we move forward with the new technology.

## Science and Mathematics Teachers' In-The-Moment Noticing: Attending to Student Thinking Within a Lesson and Beyond

Adam A. Colestock, Rosemary S. Russ

Recent reform efforts in science and mathematics education call for teachers to carefully attend and respond to their students' thinking in the classroom (Schifter, 2001; Hammer & Van Zee, 2006). To help teachers achieve this goal we must first understand their existing practices for attending to student thinking (Sherin, 2001; Levin, Hammer & Coffey, 2009). In this paper we investigate how the teachers in our sample attend to student thinking in the moments of instruction. In particular we draw on data collected with our new methodology to explore the question: "Why do particular moments of student thinking stand out as interesting to a teacher?"

For our analysis, we relied heavily on the teachers' reflections in the interviews. In particular, we used this data to explore teachers' ideas about the moments of student thinking they captured with the camera. First we reviewed the videotaped interviews and created descriptive summaries of what a teacher found interesting in each captured moment. Second, from those descriptions we identified and selected for analysis only those summaries in which the teacher discussed students' thinking as an important aspect of the reason they captured that moment. Of the 266 clip summaries we created, we identified 48% as relating to student thinking. We then looked across these summaries for evidence of why particular moments of student thinking stood out to the teacher. The four themes that emerged from the data suggest that a teacher's predictions and expectations for a particular lesson and their knowledge of their students' prior and future learning strongly influence the student thinking moments to which they attend. Below we describe these four themes, highlight how they relate to the work of teaching, and provide illustrative examples from the data.

*Theme 1: Attending to unanticipated moments of student thinking*. In planning for a lesson, teachers use their prior teaching experience, their knowledge of typical student thinking, and their understanding of this particular group of students to anticipate students' reactions to different parts of the lesson, including what students might find routine or easy and what they might find challenging or puzzling (Schoenfeld,1998). Thus teachers may have a set of expectations about what student thinking will emerge in the lesson. In our data we found that teachers often described the moments they captured as being surprising or unexpected. For example teachers sometimes focused their discussion of a moment on a particular student question or insight that they had not anticipated, but were nonetheless pleased with encountering. For example, one teacher captured a moment because a student was able to solve algebraic equations involving absolute value signs, something they had not yet learned as a class. At other times teachers described moments when students had trouble with aspects of the lesson that they had assumed would be straightforward. For example, a science teacher captured a moment in which a student has difficulty answering what she considered to be a straightforward question. In describing the moment, she explained, "I asked what the function of the cell membrane, which we've spent the last 15-20 minutes talking about… It is literally written on the slide. We've said it like ten times. And I call on Jess and she has no clue."

*Theme 2: Attending to the progress of the lesson.* Another aspect of lesson planning involves considering the pace of a lesson, the sequence of ideas, and the location of important conceptual checkpoints within the lesson (Leinhardt, 1993). When discussing why they captured particular moments of student thinking our teachers sometimes discussed the ways in which a student's idea related to the progress of the lesson. Some teachers reported specifically waiting for students to express certain ideas and capturing them because they were important benchmarks for monitoring lesson progress. For example, one science teacher captured several different moments in which her students were accomplishing her instructional goal of making the connection between the pattern recognition activity that they were engaged in and the organization of the periodic table of elements. In describing one interaction with a student she explains, "He said, 'Oh, this is just like the periodic table!' and I was like 'Ding, Ding, Ding! Yes, that is exactly what you were supposed to [figure out].'" In contrast, other teachers captured moments when the lesson was not progressing as they had hoped, such as when students struggled with a crucial part of the lesson. For example, a mathematics teacher captured a moment in which a number of students were having difficulty understanding why it wasn't necessary to write the variable $t$ as part of the answer when finding the rate of change of a parametric linear equation. She took this as evidence that she should think about doing more work to help students understand slopes and rates of change.

*Theme 3: Attending to students' prior knowledge.* In planning a lesson, teachers make assumptions about the relevant prior knowledge students possess, either from the everyday world or from previous formal teaching. In our data teachers frequently described moments as being interesting for what they revealed about a student's prior knowledge. For example, during a lecture about membrane structure one science teacher captured a moment in which a student asked what a solute was and another student provided a poor explanation. This moment of student thinking stood out to her because she was surprised that despite their previous coursework the students were unable to accurately describe a solute, "It was mind-boggling to me that they do not know what a solute is because these are kids who have come through two years of honors chemistry and physics…so I have to think about assumptions that I make about what they have learned in the past." In addition to capturing moments in which student prior knowledge was lacking or problematic, our teachers also captured moments in which students productively drew on prior knowledge by making connections with previously learned material. For example, the same science teacher captured another moment because a student applied his understanding of protein structures that he had learned in class a month and a half ago to a new context in a class research presentation about a particular disease.

*Theme 4: Attending to opportunities for future learning.* Often teachers plan to leverage the understandings that students develop in one day's lesson as starting points for future lessons in the curriculum. As a result, teachers may draw on their awareness of how ideas will be used in the future to be alert for opportunities to foreshadow or motivate future learning. Several teachers in our sample discussed the student thinking they captured in terms of its importance for future lessons. For example, one Calculus teacher captured a student question and subsequent discussion about how a particular integration technique might be similar to other methods they would encounter later in the week. He was pleased because the discussion allowed him to foreshadow upcoming lessons; he said, "it was an interesting question because we've just started talking about these solids of revolution today…but eventually we will get to what Eric was talking about." The teacher also indicated that he decided to bring the conversation to a certain point and then stop it because they would be returning to these ideas later, "If we were working on cross sections we could have spent more time with Derrick's response but as it was I thought Dylan answered the question sufficiently for today."

In this work we have examined why teachers may attend to particular aspects of student thinking in the moments of instruction and not others. In particular we discussed how their attention to events is influenced by (1) the student responses they anticipate (2) the conceptual checkpoints that help them determine when to proceed (3) the relationship of the current lesson to students' prior knowledge and (4) the opportunity of the thinking to foreshadow future learning. Furthermore, the methodology we employed to access that teacher thinking allowed us to see teachers in their own teaching situations — situations in which they have rich knowledge of the curriculum, the students' learning history, and the possible paths along which the lesson will progress. Our analysis suggested that this detailed knowledge creates expectations and predictions that substantially influence which moments stand out to teachers. Thus we suspect that other methodologies that do not allow teachers to draw on this rich set of knowledge – either because teachers are reflecting on classrooms that are not their own or because they are too far removed from the in-the-moment use of that knowledge in their own classroom – may be unable to access the kind of thinking about student thinking we describe here. In our future work we plan to continue to use this methodology to explore how teachers attend to their students' thinking during instruction and the role that this noticing plays in shaping their teaching practice.

## Supporting Video Club Conversations Using Teacher-Selected Video Clips

Melissa J. Luna, Martha Mulligan, Miriam Gamoran Sherin, Janet Walkoe

The other papers in this symposium report on the kinds of classroom moments teachers captured using wearable cameras and the reasons teachers give for capturing such moments. In this paper, we take a different approach. Specifically, we investigate using these captured moments to support conversations among teachers around their students' ideas in science and mathematics.

This work takes place in a particular context we call video clubs. A video club is a type of professional development experience in which a group of teachers watch and discuss classroom video excerpts of their instruction with a particular focus or framework in mind (Frederiksen, Sipusic, Sherin, & Wolfe, 1998). For example, discussions in a video club context can be intentionally focused on a range of issues such as discourse, student thinking, or management (Tochon, 1999). The use of video is central to the work of a video club. Video is able to capture the complexity of a classroom and meaningfully reduce that complexity by providing a record of interactions (Borko, Jacobs, Eiteljorg, & Pittman, 2008). With video, teachers do not have to respond immediately as they do when they are teaching. Therefore, watching and discussing video opens up rich opportunities for teachers to reflect upon and analyze events that occur during teaching.

Prior research on the use of video clubs demonstrated that this is an effective context for helping teachers notice and respond to students' ideas in mathematics (e.g. van Es & Sherin, 2008). This paper extends such work to include science teachers. Furthermore, the use of a new video technology—the wearable camera—inspired a slightly different video club design. Other video clubs typically ask teachers to reflect on and discuss video that researchers have selected of teachers' classrooms from footage from the back or side of the classroom (i.e., video from a researcher's point of view) (Sherin & Han, 2004). Here we instead asked teachers to reflect on and discuss classroom moments that the teachers themselves captured while wearing the Deja View or POV 1.5 camera. In doing so we had two main goals. First, we wanted to explore the ways in which clips from the wearable camera would support conversation around students' ideas. Would the fact that the video clips are exclusively from the teachers' point of view—and thus teachers are not visible—prompt consistent discussion of students' thinking? Second, we wanted to take a first step towards testing the feasibility and viability of making video clubs self-sufficient rather than relying heavily on a research team. In our experience, teachers do not usually have time to select excerpts from video of an entire class session. The wearable camera allows teachers to select the moments while they are teaching, thus removing this barrier. While we were still heavily involved in the logistics of the video clubs discussed here, we view this study as a valuable first step toward understanding what it would take for teachers to sustain a video club on their own.

The data in this study draw from our work with three separate video clubs. All video club meetings were videotaped.

(1) *High School Mathematics Video Club.* This video club consists of two experienced high school math teachers and a researcher-facilitator. Both teachers had previously participated in other video clubs to examine their students' mathematical thinking using classroom video. However, both were new users of the Deja View, which required them to capture classroom moments from their point of view for the video club discussion. Video club conversations focused on students' mathematical thinking.

(2) *Middle and High School Mathematics Video Club.* This video club consists of seven teachers with varying levels of teaching experience and a researcher-facilitator. Both the video club context and the POV 1.5 were novel for this group. At the start of the year the researcher videotaped the classroom and selected video clips for the video club discussions. Then later in the year, teacher-captured clips were used as the focus of

discussion in the video club. Video club conversations focused on classroom discourse around mathematics.

(3) *Elementary School Science Video Club.* This video club consists of four 3-5th grade teachers with varying levels of teaching experience and a researcher-facilitator. The video club context and the POV 1.5 were new for this group as well. However, this video club was different from the others in that it only used teacher-captured video clips for discussion. Video club conversations focused on students' ideas in science.

While the three video clubs reported here differed in important ways, they were similar in that they all included the use of a small wearable camera by teachers. We argue that because of its capability to capture classroom moments as they happen, this tool helped support interesting conversations in the video club as teachers reflected on and analyzed those moments. In our analysis of the video club discussions, three issues stand out as noteworthy.

First, the short time length of the videos appeared to influence the teachers' initial conversations around the video clips. Video club discussions in general focus on 5 to 10 minutes of video footage from a participating teacher's classroom. The Deja View and POV 1.5, however, capture short episodes of classroom events ranging from 30 seconds to 3 minutes. In using these shorter clips as a basis for discussion, we found that this placed greater burden on the teacher whose clip was being discussed. Essentially, these shorter clips strip away contextual details and thus require the presenting teacher to provide sufficient context in order to reconstruct the event for others. In addition, the other teachers had to articulate a range of questions as they made sense of the event. For example, after watching a 30-second clip of students from Richard's class, Richard and Nancy conversed back and forth until they arrived at a shared understanding of the moment. In the video students were discussing how they expanded the binomial $(x+5)^3$.

*Nancy:* [So] her question was, how do you multiply all three binomials at once?
*Richard:* Yeah that['s]… right.
*Nancy:* She was like… "How do you multiply three times three times three?" …
*Richard:* Right. Times three…
*Nancy:* So I think she wanted an easier way…to expand that binomial, but you haven't gotten to that. …But I don't really understand what the presenter said…
*Richard:* Well, …the first girl, her question was, … "How do you know? If you do the first and the second…or the second or the third first, which do you do?" And so what the girl in front said is that it doesn't matter because it is multiplication.

Together these two teachers carefully scrutinized the short captured clip in order to understand both what the student was asking and what the student was thinking about expanding binomials. This level of scrutiny required close attention to a classroom event on the part of both teachers.

Second, when teacher-captured clips were shown in the video club, presenting teachers had already "noticed" these moments during instruction. Thus, they typically came into the video club with an established interpretation of what had taken place in the captured moment. Perhaps using the camera during instruction heightened their noticing somewhat as well as deepened their interpretations in the moment. Interestingly, however, we found that teachers remained open to considering alternative interpretations of the events when viewed with colleagues in the video club. In fact, these new interpretations often built on the ideas the presenting teachers brought to the group. For example, upon capturing a classroom moment Nancy initially thought her students had understood the classroom talk around the equation 2x + 3:

> I think I captured it because they [the students] *did* correct it [the problem]. …the girl did say "No wouldn't that be plus 3x?" But then I think something else happened after that. I think the presenter was convinced that she didn't need to say per week, … and she did change her mind.

However, while discussing her clip in the video club she realized something different:

> Well, ...now I noticed, as I think about this a little bit more, I'm thinking how she [the student] described her situation. I think it might have been just sort of tangled in semantics. [So] is *x* the total amount of money she saved, or is *x* the amount of money she saved that week? So, right now I don't know what it was…But I think that distinction matters and so this whole conversation, I don't know that anybody was clear on that.

This differs from other video clubs because teachers in those clubs did not capture the video clips themselves, and therefore, the first explicit interpretation of the event occurred in the context of the video club. We found that having teachers bring their in-the-moment interpretations into the video club discussion, as well as their willingness to consider other interpretations, added a level of depth to unpacking the classroom event captured.

Third, at times we noted a topic of discussion that was new for video clubs. Specifically, teachers occasionally mentioned and asked each other about the reasons why they had captured particular moments with the wearable cameras. For example, in the above example, the conversation later turned to why Richard captured the clip to begin with and he responded by explaining his thinking.

> Well, it's just really about multiplying, because I was thinking…a lot of times they don't quite understand a parenthetic expression is like, you can treat it like a number. And so, in terms of applying properties, she was saying it is just multiplication, it just doesn't really matter. So I think that is probably why.

We suspect this kind of talk occurred because we asked teachers to capture specific kinds of classroom events (e.g. "Capture students' algebraic thinking." or "Capture student ideas about magnets.") while wearing the Deja View or POV 1.5. Since all teachers from one video club had the same prompt, it is not surprising they were curious of each others' reasons for capturing a particular classroom moment, and thus turned the video club discussion towards unpacking the teachers' thinking and reasoning behind the moment of capture. While we want video club discussion to focus on what students are saying and doing, we recognize that unpacking teacher thinking can be a productive focus for a video club conversation helping teachers understand their own thinking in this process.

In conclusion, video clubs that utilize clips captured by teachers in the moments of instruction do support sustainable conversations in the video club context. We have found that such conversations involve teachers in collaborative sense making often elaborating on contextual detail not apparent in the clips, being open to alternative interpretations when viewing practice, and focusing on their own thinking about why they identified a particular classroom moment as worthy of capture.

## Implications of this Methodology

Our purpose in presenting these papers as a set is to give the reader a sense of the potential of a new digital video technology – selective archiving – for the study of teacher noticing. For us, the power of these cameras lies in that they allow teachers to record particular moments of classroom interaction as the interactions unfold and without interrupting the on-going instruction. Although the full breadth of our research program with these cameras is still emerging, our use of this technology so far has provided a window into the kinds of data we might collect when working with teachers, the kinds of questions that data will allow us to answer, and the kinds of theoretical issues that data will inform.

In terms of the data we can reasonably expect to obtain with this methodology, we have seen that teachers can implement this selective archiving technology to collect moments that are meaningful to them, and that those moments can ground both personal reflections and discussions with colleagues. This process then provides us with data about the range of things that teachers notice during instruction, their own thinking about what they notice, and their colleagues' interpretations of that noticing. In terms of the questions this data can address, we see that this data may be useful for examining both population- and individual-level questions. For example, aggregating data across multiple teachers will allow us to answer questions about trends and patterns in teachers' in-the-moment noticing. In contrast, identifying moment-to-moment relationships between a single teacher's noticing as captured with the camera and that teacher's instruction may allow us to better understand the how's and why's of a successful teacher's noticing. The answers to both kinds of questions will be crucial for understanding and impacting teacher noticing. Finally, in terms of the theoretical issues our data might inform, we have found that just the possibility of capturing teachers' in-the-moment noticing has forced us to examine our underlying conceptualization of teacher noticing. For example, the methodology raises questions about whether the noticing that is conscious - and thus captured with this data - and more tacit noticing are the result of similar or distinct cognitive processes. In addition, data from teacher reflections suggests that we as researchers may need to explore the possibility that noticing is not isolated to a single event in time but is rather distributed across multiple episodes. While we do not yet have definitive answers to all these questions, our preliminary analysis of the data suggests that this new technology will allow us to make substantial contributions to the study of teacher noticing – both methodologically and theoretically - in the near future.

## References


Ball, D. L. (1997). What do students know? Facing challenges of distance, context and desire in trying to hear children. In B. J. Biddle, T. L. Good, & I. F. Goodson (Eds.), *International handbook of teachers and teaching* (Vol. II, pp. 769-818). Dordrecht: Kluwer.

Borko, H., Jacobs, J., Eiteljorg, E., Pittman, M.E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching & Teacher Education, 24*(2), 417-436.

Copeland, W. D., Birmingham, C., DeMeulle, L., D'Emidio-Caston, M., & Natal, D. (1994). Making meaning in classrooms: An investigation of cognitive processes in aspiring teachers, experienced teachers, and their peers. *American Educational Research Journal*, *31*(1), 166-196.

Frederiksen, J., Sipusic, M., Sherin, M., & Wolfe, E. (1998). Video portfolio assessment: Creating a framework for viewing the functions of teaching. *Educational Assessment, 5(4),* 225-297.

Hammer, D. & van Zee, E. (2006). *Seeing the Science in Children's Thinking.* Portsmouth, NH: Heinemann.

Leinhardt, G. (1993) On teaching. In R. Glaser (Ed.), *Advances in Instructional psychology* (Volume 4, pp. 1-54). Hillsdale, NJ: Erlbaum.

Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. *Journal of Teacher Education, 60(2),* 142-154.

Mason, J. (1998). Enabling teachers to be real teachers: Necessary levels of awareness and structure of attention. *Journal of Mathematics Teacher Education. Vol 1(3) 1998, 243-267.*

Peterson, P. L. & Clark, C. M. (1978). Teachers' reports of their cognitive processes during teaching. *American Educational Research Journal, 15*(4), 555-565.

Reich, S., Goldberg, L., & Hudek, S. (October, 2004). *Deja View Camewear Model 100*. Paper presented at the First ACM workshop on Continuous Archival and Retrieval of Personal Experiences. New York.

Schifter, D. (2001). Learning to see the invisible: What skills and knowledge are needed to engage with students' mathematical ideas? In T. Wood, B. S. Nelson, & J. Warfield (Eds.), *Beyond classical pedagogy: Teaching elementary school mathematics* (pp. 109-134).Mahwah, NJ: Lawrence Erlbaum Associates.

Schoenfeld, A. H. (1998). Toward a theory of teaching-in-context*. Issues in Education, 4*(1), 1-94.

Sherin, M. G. (2001). Developing a professional vision of classroom events. In T. Wood, B. S. Nelson, & J. Warfield (Eds.), *Beyond classical pedagogy: Teaching elementary school mathematics* (pp. 75-93). Mahwah, NJ: Lawrence Erlbaum Associates.

Sherin, M. G., & Han, S. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education, 20,*163-183.

Tochon, F. V. (1999). *Video study groups for education, professional development and change*. Madison, WI: Atwood Publishing.

van Es, E.A. & Sherin, M.G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education, 24*, 244-276.

V.I.O., Inc. (2009). POV. 1.5. Retrieved September 21, 2009 from http://www.vio-pov.com.

# Software-Based Scaffolding: Supporting the Development of Knowledge Building Discourse in Online Courses

Nobuko Fujita, Christopher Teplovs
OISE/University of Toronto, 252 Bloor Street West, Toronto, ON, M5S 1V6,
nobuko.fujita@gmail.com, christopher.teplovs@gmail.com

**Abstract:** This design-based research study investigated instructional scaffolding for knowledge building discourse among participants (n=17, n=20) in two online graduate courses. In particular, designs of software-based scaffolding as found in web-based Knowledge Forum's scaffold support feature were refined. Analyses of the student discourse data suggests that Knowledge Forum's scaffold supports offer a promising avenue for future design innovations to encourage knowledge building discourse. Results show that students increasingly used the scaffolds to focus their reading and writing of notes over iterations of the study. The proportion of scaffolds for knowledge building discourse increased during each iteration with a corresponding decrease in the proportion of scaffolds for expressing an opinion in the second iteration. Finally, notes with scaffolds contained significantly more words than notes without scaffolds, suggesting that scaffolds promoted more student reflectivity. Implications for formative assessment of student learning and knowledge building are discussed.

## Introduction

In recent years, opportunities to enroll in online courses have grown substantially. During the fall 2004 term, 2.3 million higher education students were taking at least one online course in the United States; one year later, during the fall 2005 term, this figure increased to nearly 3.2 million students (Allen & Seaman, 2006). The primary mode of online course delivery at nearly 90% of U.S. higher education institutions is asynchronous computer-mediated communication (CMC) (Greene, 2005). Asynchronous CMC typically uses a web-based computer conferencing technology (e.g., Blackboard, Drupal, Knowledge Forum, Moodle, etc.) to connect distributed participants in a networked learning environment. Like distance education, CMC frees students from time and space constraints; like face-to-face instruction, CMC affords interactivity (Kaye, 1989). Ideally, CMC is used to create a "the kind of learning community that can arise in a good graduate seminar" (Hiltz, 1998). In such educational settings, students may be able to engage in complex learning where they share, question and revise their ideas to deepen understanding and build knowledge.

Yet CMC is a lean medium for social presence crucial to perceptions of cooperation and trust necessary to build a learning community (Palloff & Pratt, 1999; Preece, 2000). Thus, widely accepted models for teaching in online learning communities focus first on establishing social connections between participants. For example, Salmon's (2000) five-stage model for "e-moderating" describes the progression of students as they move through the beginning stages of access and motivation, to online socialization, information exchange and knowledge construction, and ultimately to development. This model highlights the importance of the instructor's role in the earlier stages of creating an online learning community. Once a sense of community is established, the instructor shifts the locus of control to the students, who are then able to engage in active learning and collaboration (Palloff & Pratt, 1999).

This study departs from previous studies by focusing on higher goals for social interaction in online learning communities. Research indicates that critical thinking and knowledge construction rarely occur in these educational settings (Garrison, Anderson & Archer, 2001; Gunawardena, Lowe, & Anderson, 1997; Kanuka & Anderson, 1998). Previous studies on scaffolding dialogue for knowledge building have explored how certain critical thinking types increase or decrease the length of discussion threads (e.g., Sorenson & Takle, 2005). However this study takes an approach consistent with a knowledge building pedagogy (Scardamalia & Bereiter, 2003, 2006), rather than one informed by a critical thinking model. While many online studies have examined social interaction and collaboration, less is known about how instructors may move students to the next step and foster knowledge building discourse. Therefore, the purpose of this study is to investigate what kinds of scaffolding embedded in the software and course activity structures can support the development of knowledge building discourse in online graduate courses. As well, we argue that scaffolding is closely linked to formative assessment (c.f., Shepard, 2005) and that can activate students to take ownership over their learning (Black & Wiliam, 2009) in a way that is aligned with the knowledge building pedagogy.

## Theoretical Framework

Social constructivist pedagogy emphasizes the importance of discourse in fostering deep learning, and the importance of tools in mediating knowledge construction (Hmelo-Silver, 2003; Palincsar, 1998). Knowledge building is defined as "the production and continual improvement of ideas of value to a community" (Scardamalia & Bereiter, 2003, p. 1370). It is a constructivist approach that centers the curriculum on big ideas and accords students with high levels of agency in working with those ideas (Scardamalia & Bereiter, 2008).

Knowledge Forum, an extension of CSILE (Computer Supported Intentional Learning Environment) software, is designed to support knowledge building. Students work in virtual spaces or "views" to develop their ideas, represented as "notes." Knowledge Forum offers sophisticated features not available in other conferencing technologies including "scaffold supports" (labels of thinking types), "rise above" (summary note), and a capacity to connect ideas through links between notes in different views. These features provide means to overcome the chronological sequence of threaded discussion, in which important ideas may be lost. In addition, Knowledge Forum facilitates the collection of data that are amenable to analysis with a variety of assessment tools. These include behavioral and interaction analyses (Burtis, 1998), traces of vocabulary development (Hewitt, 1999), social network analysis (Teplovs, Donoahue, Scardamalia, & Philip, 2007), and semantic analysis (Fujita & Teplovs, 2009). Such assessments provide feedback to help students make progress in their discourse.

Many studies on knowledge building discourse have examined elementary science classrooms (e.g., Hakkarainen & Palonen, 2003; Bereiter, Scardamalia, Cassells, & Hewitt, 1997). These show that CSILE/Knowledge Forum can support improved learning and knowledge building over time (Scardamalia & Bereiter, 1994; Zhang, Scardamalia, Reeve & Messina, 2009). However, few studies have explored how to support the development of knowledge building discourse in exclusively online higher education courses where the participants are distributed across time and space and instructional scaffolding for knowledge building is not provided off-line. The knowledge building teacher's role in this setting is one that helps students move beyond the formation of social ties to the creation of knowledge. This study offers a unique perspective by refining designs of software-based scaffolding as found in Knowledge Forum's scaffold support feature to foster knowledge building discourse in online graduate courses.

## Methods

Using a design research methodology (Collins, Joseph, & Bielaczyc, 2004), modifications to the selection of Knowledge Forum's (KF) scaffolds were made within and across two 13-week online graduate courses, which comprised two iterations of the study (Table 1).

Table 1: Design Changes to Knowledge Forum's Scaffold Across Two Online Graduate Courses

| | Iteration 1 | Iteration 2 |
|---|---|---|
| Modification | Educational Applications of Computer-Mediated Communication | Constructive Learning and Design of Online Learning Environments |
| Knowledge Forum Scaffolds | Theory Building<br>Opinion<br>Idea Improvement<br>Feedback | Theory Building<br>Opinion<br>Idea Advancement |

KF scaffold supports are typically used as sentence openers that students use while composing notes in the database. They function in similar ways as prompts in social networking applications like Facebook (What's on your mind?) or Twitter (What are you doing?) by placing yellow highlights of thinking types into the text that bracket segments of body text in notes. Scaffold supports go further than just prompting users at a generic level. They are customizable by facilitators and, in some cases, students. The two sets of scaffold supports that are supplied by default ("Theory Building" and "Opinion") are more specific to knowledge building discourse and argument than the more generic prompts from Facebook and Twitter. Scaffold supports are not dissimilar to "tags" that are common in other Web 2.0 environments such as the collaborative bookmarking application known as "delicious". An important difference between the simple tagging available through such sites and the scaffold supports in Knowledge Forum is the ability to tag specific parts of posts and thereby indicate with relatively high specificity and accuracy that part of the text that is being tagged. Moreover, advanced search capabilities in Knowledge Forum allow users to search for specified text within a specified scaffold support (e.g. find all the "My Theory" supports that contain the word "constructivism").

In Iteration 1, only Theory Building and Opinion scaffolds built in to Knowledge Forum were available to students at the beginning of the course. Later, in week 9 of 13 weeks, students took on the responsibility for designing the customized Idea Improvement scaffolds as part of their discussion leadership (Table 2). These scaffolds emphasize the knowledge building principle of improvable ideas key to knowledge building discourse. In addition the researcher introduced the customized Feedback scaffolds in week 10 to help students provide constructive feedback to each other.

Changes to the scaffolds from Iteration 1 to 2 resulted from analysis of the students' learning journals. The design researcher and the instructor addressed students' concerns for the constraint scaffolds imposed on creative thinking and interpersonal relationships were addressed by combining the Idea Improvement and Feedback scaffolds into Idea Advancement scaffolds in Iteration 2.

Table 2: Knowledge Forum Scaffolds and Scaffold Supports Used in Iteration 1 and 2

| | Scaffolds | | | |
|---|---|---|---|---|
| Scaffold Supports | Theory Building | Opinion | IDEA IMPROVEMENT (all caps in original) | Idea Advancement |
| | My Theory | Opinion | IDEA ADVANCEMENT | Current statement of idea |
| | I need to understand | Different opinion | WHAT DO WE NEED THIS IDEA FOR | How idea is useful |
| | New information | Reason | PROBLEM/QUESTION | How idea could be advanced |
| | This theory cannot explain | Elaboration | | Problems/limitations |
| | A better theory | Evidence | | How could we test X? |
| | Putting our knowledge together | Example | | This idea fits with |

## Data sources

Student discourse, online questionnaire responses, pre- and post-course assignments, and learning journals were collected from participants (n=17, n=20) in two 13-week graduate courses taught entirely online using Knowledge Forum (version 4.5.3). This study focuses on findings from the analysis of KF scaffold use in the discourse data.

## Results

Using log file data accessed via the Analytic Toolkit built in to Knowledge Forum, this study found patterns in the students' use of KF scaffolds that affected the development of knowledge building discourse. First, there was an overall increase in the total number of scaffolds used from Iteration 1 to Iteration 2, as shown in Table 3:

Table 3: Frequencies of Knowledge Forum Scaffolds Used in two Online Graduate Courses

| Course | Theory Building | | Opinion | | IDEA IMPROVEMENT/ Idea Advancement | | Total Number of Scaffolds | | Total Number of Student Notes | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *f* | % | *f* | % | *f* | % | *f* | % | *f* | % |
| Course 1 | 128 | 42% | 91 | 30% | 87 | 28% | 306 | 100% | 907 | 100% |
| Course 2 | 167 | 36% | 200 | 43% | 98 | 21% | 465 | 100% | 1058 | 100% |

Course 2 students used more total scaffolds than Course 1 students. Course 2 students also used more Opinion than the Theory Building or Idea Improvement scaffolds than Course 1 students. A Pearson chi square test showed a statistically significant difference between the two Iterations, $\chi^2$ (2, N=771) = 14.46, p<.001.

### Relationship Between Use of Scaffolds and Types of Scaffolds Used

To understand whether the types of scaffolds used in student notes changed over time within each course, the frequencies of different types of KF scaffolds used were calculated for the first third and last third of the course, excepting the first week and last week. These weeks were omitted because they were used as introductory and evaluation sessions. When calculated, this analysis showed that Course 1 students used all of the scaffold types more frequently in the last third of the course compared to the first third (Figure 1).

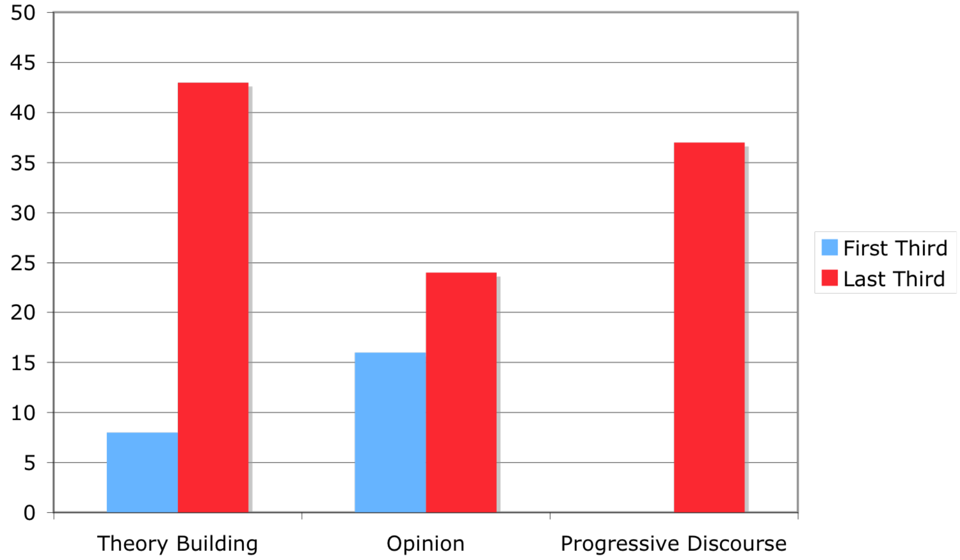



Figure 1. Types of Knowledge Forum scaffolds used by students in Course 1.

In comparison, all three types of scaffolds were available from the beginning of Course 2. Promisingly, the students' use of Theory Building and Progressive Discourse (Idea Improvement and Idea Advancement) scaffolds increased and their use of Opinion scaffolds decreased from the first third to the last third of the course (Figure 2).

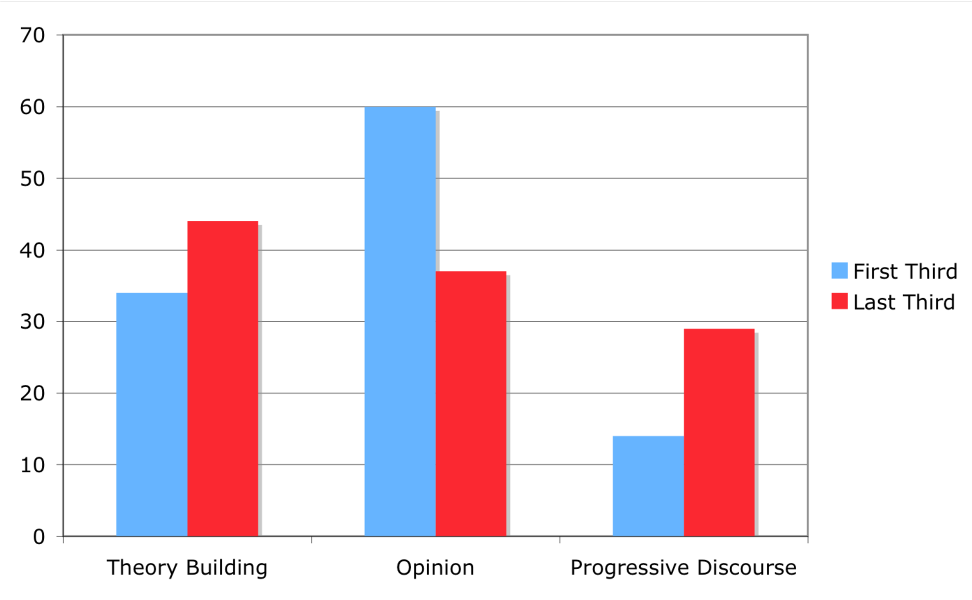



Figure 2. Types of Knowledge Forum scaffolds used by students in Course 2.

### Relationship Between Use of Scaffolds and Length of Notes

To understand the relationship between the use of KF scaffolds and length of notes, the mean word counts of student notes with scaffolds and without scaffolds were compared. A paired samples t-test found that notes with scaffolds contained significantly more words than notes without scaffolds, $t\,(33)=3.626$, $p<.001$.

## Discussion

This design-based research study refined designs of software-based scaffolding as found in Knowledge Forum's scaffold support feature to foster knowledge building discourse in two online graduate courses. KF scaffolds not only metacognitively prompt students to focus their reading and writing of notes, but also allow them to tag relevant parts of notes and search for ideas to improve them. Such a feature offers students, not just teachers, the opportunity to take responsibility for accessing support as needed, customizing supports for local needs, and obtaining some formative assessment of the progress of the unfolding discourse to deepen their learning. Analyses of the student discourse data suggest that KF scaffold supports offer a promising avenue for future design innovations to encourage knowledge building discourse.

Finding that students used scaffolds for expressing opinions aimed at argument rather than scaffolds to support knowledge building discourse aimed at explanation was disappointing, but not surprising given the challenge of engaging students in this kind of discourse to promote complex learning online. Scaffolds can guide students in complex learning by structuring tasks, but they can also problematize tasks by making them more complex in the short term (Reiser, 2005).

Over time within course discussion spaces, however, students increasingly used customized scaffolds to support development of knowledge building discourse. In Iteration 2, students not only increased their use of scaffolds for knowledge building, but they also decreased their use of scaffolds for expressing opinions. This is a positive finding since using the Theory Building and Progressive Discourse scaffolds require students to take a more active role in regulating their own learning as well as providing constructive feedback to peers.

From a knowledge building perspective, a characteristic of a mature producer of knowledge is "disciplined creativity" (Scardamalia & Bereiter, 2003). KF scaffolds help students to organize their own thinking, writing, and reading of other's notes. This may promote disciplined practices for knowledge building discourse. Unfortunately, some students in this study (35% in Iteration 1; 25% in Iteration 2) did not like the "disciplined" aspect of creative knowledge work that scaffolds were designed to support and identified constraints scaffolds posed for creative thinking and identified usability issues.

Previous researchers (e.g., Hara, Bonk, & Angeli, 2000; Schrire, 2006) have suggested that the length of messages is one sign of the depth of online student interaction and reflection on course readings. This study found that student notes with KF scaffolds are longer than notes without them. Students might be reflecting more when they compose longer notes with scaffolds and thus more effectively engage in knowledge building discourse to advance knowledge in their online learning community.

## Conclusion

In this paper, we explored designs of software-based scaffolds and course activity structures that encourage their use to support the development of knowledge building discourse in online graduate courses. Through analyses of students' scaffold use, this study found the following trends in the students' use of the Knowledge Forum's scaffold support feature:

- an increase in the use of scaffolds over iterations of the study
- the proportion of scaffolds for knowledge building discourse increased during each iteration
- the proportion of Opinion scaffolds for argument decreased during the final iteration
- notes with scaffolds contained significantly more words than notes without scaffolds

The students appear to be taking greater responsibility for their own learning by customizing and using the KF scaffolds to make their online course contributions more meaningful. Qualitative analyses reveal that the graduate students use KF scaffolds to work towards common understanding in discourse, but need encouragement to expand the factual base and to test ideas to build knowledge (Fujita, 2009). Formative assessment that is embedded in the software and tightly coupled to the learning (Collins & Halverson, 2009) may help students overcome such barriers in knowledge building discourse. To this end, Knowledge Forum's assessment tools offer ongoing and formative feedback to students as well as teachers and researchers to enable them to deepen understanding of the unfolding discourse in the online learning community (Teplovs et al., 2007).

## References


Allen, I. E., & Seaman, J. (2006). *Making the grade: Online education in the United States*. Needham, MA: Alfred P. Sloan Foundation

Bereiter, C., Scardamalia, M., Cassells, C., & Hewitt, J. G. (1997). Postmodernism, knowledge building, and elementary science. *Elementary School Journal, 97*(4), 329-340.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31.

Burtis, J. (1998). *Analytic Toolkit for Knowledge Forum.* Centre for Applied Cognitive Science, The Ontario Institute for Studies in Education/University of Toronto.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences, 13*(1), 15-42

Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology*. New York, NY: Teachers College Press.

Fujita, N. (2009). *Group processes supporting the development of progressive discourse in online graduate courses.* Unpublished Doctoral Dissertation, University of Toronto, Toronto, ON.

Fujita, N. & Teplovs, C. (2009, June). Automating analysis of collaborative discourse: Identifying idea clusters. In C. O'Malley, D. Suthers, P. Reimann & A. Dimitracopoulou (Eds.), *Proceedings of the Computer Supported Collaborative Learning 2009 – Volume 2* (pp.162-169*)*. Rhodes, Greece: International Society of the Learning Sciences, Inc.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence and computer conferencing in distance education. *The American Journal of Distance Education, 15*(1), 7-23.

Greene, B. R. (2005). *Distance education at higher education institutions 2000-01 (PEQIS 13)* (No. 2005118). Washington, DC: National Center for Education Statistics.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*(4), 397-431.

Hakkarainen, K., & Palonen, T. (2003). Patterns of female and male students' participation in peer interaction in computer-supported learning. *Computers & Education, 40*(4), 327-342.

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science, 28*, 115-152.

Hewitt, J. (1999). *The Vocabulary Analyser for Knowledge Forum*. Centre for Applied Cognitive Science, The Ontario Institute for Studies in Education/University of Toronto.

Hiltz, S. R. Hiltz, S. R. (1998). Collaborative learning in asynchronous learning networks: Building learning communities [Electronic Version]. Invited address at "Web98". Retrieved February 27, 2010 from http://web.njit.edu/~hiltz/collaborative_learning_in_asynch.htm.

Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education, 41*(4), 397-420.

Kanuka, H., & Anderson, T. (1998). Online social interchange, discord, and knowledge construction. *Journal of Distance Education, 13*(1), 57-74.

Kaye, A. (1989). Computer-mediated communication and distance education. In R. Mason & A. Kaye (Eds.), *Mindweave: Communication, computers and distance education* (pp. 3-21). Oxford: Pergamon Press.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 49*, 345-375.

Palloff, R. M., & Pratt, K. (1999). *Building learning communities in cyberspace: Effective strategies for the online classroom*. San Francisco: Jossey-Bass Publishers.

Preece, J. (2000). *Online communities: Designing usability, supporting sociability*. Chicester: John Wiley & Sons, Ltd.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing problem work. *The Journal of the Learning Sciences, 13*(3), 273-304.

Salmon, G. (2000). *E-Moderating: The key to teaching and learning online*. London, UK: Kogan Page.

Scardamalia, M., & Bereiter, C. (2003). Knowledge Building. In *Encyclopedia of Education* (2nd ed.). New York: Macmillan Reference, USA.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences*. New York: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2007). Fostering communities of learners and knowledge building: An interrupted dialogue. In J. C. Campione, K. E. Metz & A. S. Palincsar (Eds.), *Children's learning in laboratory and classroom contexts: Essays in honor of Ann Brown* (pp. 197-212). Mahwah, NJ: Erlbaum.

Schrire, S. (2006). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers & Education, 46*, 49-70.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership, 63*(3), 66-70.

Sorensen, E. K., & Takle, E. K. (2005). Investigating knowledge building dialogues in networked communities of practice. A collaborative learning endeavor across cultures. *Interactive Educational Multimedia, 10*, 50-60.

Teplovs, C., Donoahue, Z., Scardamalia, M., and Philip, D. (2007). Tools for concurrent, embedded, and transformative assessment of knowledge building processes and progress. In C. A. Chinn, G. Erkens, and S. Puntambekar (Eds.) *Proceedings of the 8th International Conference on Computer Supported Collaborative Learning.* New Brunswick, New Jersey: International Society of the Learning Sciences, 721-723.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge building communities. *The Journal of the Learning Sciences, 18*(1), 7-44.

# Seeing Algebraic Thinking in the Classroom: A Study of Teachers' Conceptualizations of Algebra

Janet Walkoe, Northwestern University, 2010 Campus Drive, Evanston, IL, 60208,
Walkoe@northwestern.edu

**Abstract:** Ideas about how students learn algebra have shifted over the years. Reform curricula have been developed to address these changes. Teachers' conceptualizations of algebra play a large role in how they view and implement reform curricula. This paper explores teachers' conceptualizations of algebra and investigates how they draw on their conceptualizations in the moments of instruction through the use of new methods that incorporate teacher-selected classroom video clips. Data revealed that teachers' verbal descriptions of student algebraic thinking did not always align with moments of instruction they selected as providing evidence of student algebraic thinking.

## Introduction

Algebra is seen as a critical subject matter for students. In particular, algebra has been identified as the "gatekeeper" to advanced mathematics. Students need to have taken algebra in order to have access to upper-level mathematics courses. Such courses can then lead to further opportunities such as admissions to colleges and universities. Despite its importance, many students struggle in algebra classes (Kieran, 1992). This has prompted inquiry into the question of how to effectively support student learning of algebra.

Ideas about how students best learn algebra have recently shifted. Up until the last decade or so, algebra classes emphasized procedural activities such as symbol manipulation over more conceptual activities. In addition, procedural activities were generally separated from the conceptual sense-making activities (Kieran, 2004). This was in line with the "symbol precedence" view of algebra that most textbooks presented. According to Nathan and Koedinger (2000a) the belief underlying this view is that "arithmetic reasoning strictly precedes algebraic reasoning, and symbolic problem-solving develops prior to verbal reasoning" (Nathan & Koedinger, 2000a, p. 209). In contrast to the symbol precedence view, more popular today is an emphasis on the conceptual nature of algebra and of the importance of connecting symbolic and conceptual learning (Kaput, 2008). In response to this shift, reform curricula have been developed that are more in line with current views of algebra and algebraic thinking. However, research has shown that it can be difficult for teachers to teach reform curricula in the ways that they were intended (Cohen, 1990). One reason is that teachers' ideas about algebra may not align with the goals of reform. In this project, I investigate new methods for exploring teachers' conceptualizations of algebra and algebraic thinking and how these conceptualizations can inform practice.

## Teachers' Conceptualizations of Algebra

There is much discussion in the literature about what constitutes algebra and algebraic activity (Battista and Borrow, 1998; Kaput, 1998; Kieran, 1996; NCTM, 1989; Usiskin, 1988). Kieran (1996) identified three forms of algebraic activity: generational, transformational, and global/meta-level. Generational activities include generating equations and expressions from data, situations, patterns of numbers, or number sequences. Transformational activity includes manipulating symbols following the rules of algebra. This includes typical classroom activities such as simplifying an expression, expanding a polynomial, and solving equations. Global/ meta-level activity includes using algebra to solve problems. This includes modeling, problem solving, providing evidence, and generalizing.

Traditionally, mathematics classes tend to give precedence to transformational activities. Such classrooms rely on highly procedural activities such as symbolic manipulation and focus little attention on more conceptual understanding that includes global/meta-level activities (Kieran, 2004). Recent shifts in the teaching of algebra recommend that transformational activities no longer be distinct from conceptual understandings. Therefore reform curricula have moved toward a balance among the forms of algebraic activity, emphasizing conceptual activities such as generalization, argumentation, and justification (Kaput, 1998; Carpenter, Franke & Levi, 2003; Kieran, 2004).

Teaching algebra from a more balanced perspective can be difficult for teachers for a number of reasons. First, while learning algebra themselves, most teachers experienced classes that emphasized procedures and symbol manipulation. As a result of this "dilemma of experience" (Doerr, 2004), teachers tend to hold on to this view and focus on procedural knowledge in their classrooms (Menzel & Clarke, 1998; Nathan & Koedinger, 2000b; Stephens, 2008). Second, because traditional methods are what we find in most algebra textbooks today, many teachers have not had the opportunity to explore a more conceptual approach to teaching algebra. Even when given reform materials, teachers are often asked to implement them without much guidance and they may end up interpreting the new materials in terms of their familiar approaches to the topic (Sherin, 2002). This is reminiscent of Cohen's (1990) description of Ms. Oublier, who interprets a reform curriculum in light of her traditional teaching methods.

## Conceptualizations of Algebra in Practice

This paper seeks to understand teachers' conceptualizations of algebra and the relationship between these conceptualizations and teachers' instructional practices. In particular, the goal is to investigate the conceptualizations that teachers draw on while teaching. This is somewhat of a challenge, however. One way to explore a teacher's conceptualizations is through interviews where teachers are asked to explain their ideas. We know, however, that a teacher's espoused beliefs do not always align with their teaching practice (Cooney, 1985; Cohen, 1990). For example, Cooney (1985) describes an example of a teacher who, when interviewed, said he believed problem solving was essential to mathematics learning yet, due to a number of factors, did not incorporate much problem solving in his instruction. In order to explore the relationship between teachers' conceptualizations of algebra and their practice, we need to look at both espoused views and views as enacted.

One way to gain information about a teacher's enacted views is to observe her class instruction. However, one problem with observing a class and making inferences is that we do not know if our interpretation is the same as the teacher's. We might interview a teacher after class to try to access her thinking during instruction, but she may not remember what she was thinking in the moment of instruction. Many times teachers recall general things about a class based simply on overall feelings about how the instruction took place (Rosean, Lundeberg, Cooper, Fritzen, Terpstra, 2008). A related approach to exploring a teacher's practice is through the use of video.

## Video to Explore Teacher Thinking in the Moments of Instruction

Video has been a valuable tool used to examine a teacher's practice (Lampert & Ball, 1998; Sherin & van Es, 2005; Sherin, 2007). Video offers the advantage of allowing observers to watch events as they took place instead of relying on memory or observation notes. This allows teachers the ability to "get back into" the moment of instruction. However, while the use of classroom video gives us valuable insights, it raises questions about whether a teacher's attention to student thinking after the fact is different from her attention in the moment of instruction. Relying on classroom video may draw a teacher's attention to things that they did not notice in the moments of instruction. Ainley and Luntley (2007) used video as a way to access teachers' in-the-moment reasoning. As they observed, however, "in some interviews, the effect of watching the video was to draw attention to aspects of the classroom that the teacher had previously been unaware of, rather than to help recall their perceptions of the lesson" (p. 9). So we cannot always be sure an interview based on classroom video taps into what a teacher was thinking while she was teaching.

To address this issue, new methods have been explored that make use of a camera that allows a teacher to record moments that catch her attention while she is teaching (Sherin, Russ, Sherin, & Colestock, 2008). This allows access to in-the-moment noticing that was not available before. Sherin, Russ, Sherin, & Colestock (2008) piloted the Camwear 100 (Reich, Goldberg, Hudek, 2004) that allows teachers to capture short video clips while they are teaching. It involves a small camera a teacher can wear on her person while she is teaching. When the teacher presses the record button, the camera captures the previous 30 seconds and saves it to a small digital video file to be viewed later. Using this technology allows me to study teachers' in-the-moment thinking in a way that has only recently become possible. By asking a teacher to press the button while teaching, I can identify moments the teacher was attending to in the midst of instruction and can save these moments to discuss later.

One question that might arise with this method is whether using this technology actually allows access to a teacher's in-the-moment noticing or whether it instead prompts an ad-hoc development of the vision of what is going on at the time. Sherin, Russ, and Colestock (in press) provide evidence that interviewing a teacher about the clips they captured using this technology can in fact bring teachers back to the moment and does help them talk about their thinking during instruction.

In this study, I make use of this methodology to create a more complete picture of how teachers' conceptualizations of algebra inform a teacher's practice. By using methods that allow access to teacher-selected classroom video, I can identify the moments a teacher notices while teaching. I use information about what teachers notice and how they reflect on these moments, in addition to classroom observations, to help better explore their ideas about algebra and how they draw on these ideas in the moments of instruction.

## Methods

This study explored high school teachers' ideas about student algebraic thinking. Four high school algebra teachers who teach in a local urban district volunteered to participate in this study. Two of the teachers, Dana and Maggie, taught from a traditional curriculum. Maggie taught ninth grade algebra and Dana taught eleventh grade algebra. Their teaching experience was four and nine years respectively. Nancy and Richard taught from a reform curriculum. Similar to the other pair, Nancy taught ninth grade algebra and Richard taught eleventh grade algebra. Their teaching experience was thirteen and seven years respectively. The traditional curriculum espoused a symbol precedence view whereas the reform curriculum placed a focus on student sense-making and emphasized the conceptual nature of algebra. The data consists of interviews, class observations, and selected excerpts of instruction chosen by the teachers.

### The study protocol and instrumentation

#### Pre-observation interview

Each teacher participated in a pre-observation interview designed to investigate his or her ideas about algebra and algebraic thinking. The pre-interview had two sections, the *general question* section and the *concrete student response* section. Each section of the interview lasted around 20 minutes. The general question section consisted of questions about the teacher's ideas about algebra. I asked direct questions to try to get a sense of the teacher's ideas about algebraic thinking. An example of a question from this phase was, "What do you mean when you say that you want to get students to think algebraically?" In the student response section, I showed the teacher sample student responses to problems and asked the teacher whether he or she thought the student was thinking algebraically based on the solution presented.

#### Classroom observation

The Camwear 100 is a small, portable camera developed by Dejaview (Reich, Goldberg, & Hudek, 2004). When the camera is on it continually records and erases footage in a loop mode until the "record" button is pushed. When the user presses the record button, the camera saves the preceding 30 seconds. These clips can be downloaded, saved and viewed on a computer. I videotaped each classroom session from the back of the room. During the session, I asked teachers to wear the camera while teaching. I asked the teachers to press the record button whenever they saw evidence that their students were thinking algebraically.

#### Post-observation interview

After the classroom observation, I interviewed the teachers and asked them about each clip they captured and why they thought their students were thinking algebraically in each clip. I wanted to understand the nature of student thinking the teachers were paying attention to while they were teaching. Although I cannot claim that the camera is a direct window into the teachers' thinking, I can use the clips the teacher captured as a slice of evidence of what they were paying attention to while they were teaching. My intention for the post-interview was to help clarify or provide additional insight into the moments the teacher captured and why those moments stood out to the teacher. Furthermore, it would permit triangulation with the pre-interviews in order to gain a better understanding of the teachers' ideas about algebraic thinking.

### Data analysis

In the first stage the type of activity the students were engaged in during the clips was coded. This was done by viewing the clip in the context of the whole class video that had been taken from the back of the room. This allowed me to see the context of each clip in the larger landscape of the class activities. There were two broad codes, procedural manipulation and conceptual. The procedural manipulation code was applied when the student or students in the clip were in the midst of solving a symbolic algebra problem or discussing how they had solved the problem. An example is when Maggie asked her students to compute the side length of a triangle using the Law of Sines. The conceptual code was applied when the student or students in the clip were involved in a larger sense-making activity. An example is when the students in Richard's class extended a problem he asked them to do and discussed whether the procedures could be generalized from two dimensions to three or more dimensions and if so, what it might look like.

Then the post interview in which each teacher discussed the reasons for capturing each clip was reviewed. In this first pass, I was able to distinguish between the clips the teachers had a reason for capturing and those they did not. It turned out that in a few cases, upon further reflection of the clip, the teacher decided that he or she would not consider the thinking the students were engaged in during the clip to be algebraic so those clips were removed from the next stage.

In the third stage the moments the teacher captured and the reasons they gave for capturing that moment were coded. The unit of analysis was the teacher's reason for capturing the clip. I was interested in why the teacher thought the student or students in the moment they captured were thinking algebraically. I engaged in the process of open coding where I looked at some of the teachers' responses and came up with some categories. I continued with these categories and the main reasons for identifying algebraic thinking that emerged were students (1) making connections, (2) generalizing, (3) justifying, articulating or testing, (4) understanding rules or properties of algebraic structures, (5) manipulating symbols or solving an equation, and (6) knowing or following rules of algebra. The first four were considered conceptual reasons and the last two were considered procedural reasons. Each teacher's response was assigned one code. After coding the teachers' reasons for capturing clips, I made a table containing a summary of the information. I looked for patterns in each teacher's responses. I also looked for patterns across teachers.

In the fourth stage I reviewed the pre-interviews. I was curious about whether or how ideas that emerged in the pre-interviews influenced the moments the teacher captured and how the teacher talked about those moments. I was interested in each teacher's ideas about algebra, goals for students, and feelings about their curriculum that emerged in the interviews. The analysis of the pre-interviews took place in three steps. In the first step, I segmented the transcripts and identified which portions dealt with ideas about algebra, goals for students, and thoughts on the curriculum. In the second phase, I looked at all the segments having to do with each of the three categories and I identified the central idea the teacher discussed in those segments. I did this iteratively and came up with a set of central themes the teacher discussed. In the last stage, I identified whether the themes corresponded with procedural or conceptual ideas about algebra.

Finally, I reviewed the class videos and created a summary. In the summary, I included three aspects of the class and the structure of the lesson. They were the way the class was organized, the mathematical topic that was covered, and the types of problems the students were asked to work on.

## Results and Discussion

I found that two teachers' espoused ideas about algebra aligned with the clips they captured in the classroom, therefore adding evidence to their conceptualizations of algebra as described in their interviews. One of the teachers, Maggie, held a symbolic view of algebra and this was consistent in her pre-interview and in the clips she captured and the way she discussed them in the post-interview. For example, in one clip, while Maggie was presenting a problem on the overhead a student volunteered a solution to the problem. Maggie captured the clip and when asked why she captured the clip in the post-interview she responded, "(the student) was doing basic problem solving equations. They were solving for x. Solving for the variable."

Another teacher, Nancy held a conceptual view of algebra and this was consistent with the way she discussed algebra in the pre-interview and in the clips she captured and discussed in the post-interview. In one of Nancy's captured clips a student was explaining why he thought the quadratic function the class was graphing would not "bounce" at zero. When asked about the clip in the post-interview said she thought the student was thinking algebraically because he understood the properties of parabolic functions and understood how the graph should look. He understood how symmetry was connected to the shape of the graph and saw how the graph related to the equation.

While Maggie and Nancy captured clips that supported the ideas that emerge in their pre-interviews, the other two teachers were less consistent in the way they talked about algebra in the pre-interview and the clips they captured while teaching.

## Dana

Dana taught eleventh grade algebra from a traditional curriculum. She talked about algebra from both a symbol manipulation and conceptual point of view in the pre- interview. When she discussed the clips she captured in class, however, she discussed 8 of the 11 clips with a focus on symbol manipulation.

The aspect of student thinking that stood out to Dana was the students' facility with the symbols and rules. Even though this would appear to be evidence that Dana's conceptualizations of algebra are dominated by algebraic procedures and symbol manipulation, her pre-interview told a different story.

The primary theme that emerged in Dana's pre-interview was that she believed there were two types of algebra, school algebra and more "sophisticated algebra" that is not taught in schools. When presented with a student's less traditional, more conceptual solution in the pre-interview, she commented, "(it's) logical reasoning. And it's algebra but it isn't the kind of thing that gets taught… explicitly taught." In response to another student's similar solution she said, "(the solution) is more logic than…official algebra. But I still think they're a thinking person's approach to the problem… yeah, I think that's algebra. It's just not the rote algebra we have taught." She continued, "We teach them algorithms so they can develop that. I think (the solution the student had is) really sophisticated. But this (pointing to a more symbolic answer) is what I teach." So Dana is capable of recognizing two types of algebraic thinking- procedural and conceptual algebraic thinking. In the classroom, however, she tends to focus on the procedural aspects of algebraic thinking. For example, she captured a moment in her class in which a couple of students were trying to explain why negative numbers are not in the domain of the log function. In the post-interview she said that she thought the students were thinking algebraically because they were "defining how you do algebra" and they were discussing the rules of the game. Instead of focusing on the conceptual aspects of the conversation, she focused on the students' knowledge of the rules. In another clip she captured a student answering a problem she posed on the overhead projector. The problem had more variables than the previous problems had. Dana said the student was thinking algebraically because she was able to manipulate the symbols to solve the more difficult type of problem. She commented:

> …(some students) may have said '… there are so many pieces here, how do I plug this into where it goes and this into where it goes? I've got four pieces and I've only got three slots, what am I going to do?' So I was relieved to see that most of (the students in this class) just looked at it and went, 'oh', and just did it. And saw that it was actually a simple problem. It's a little bit rote, I mean I've got- here's this equation and here's this equation and here's how you transform them. It may not be super meaningful, but… but it's algebra *(laughs)*.

Her focus was primarily on the students' realization that they could manipulate this problem as they had done with others in the past. She did not focus on the student's conceptual understanding of the problem. Overall, Dana tended to focus on students' knowledge and manipulation of the rules of algebra during instruction.

Dana exhibited dual views of algebraic thinking in the pre-interview. One is more conceptual and she believes is not the type they explicitly teach in schools. The other is more procedural and is emphasized in her curriculum. Even though she is capable of recognizing the more conceptual algebraic thinking she did not use that to inform the moments she captured in class or the way she discussed those moments. She could have discussed the same clips she captured in a more conceptual way. In the next case, we will see an example of a teacher who, similar to Dana, primarily captured clips where students were engaged in procedural activity but who discussed them in a more conceptual way.

## Richard

Richard taught eleventh grade algebra from a reform curriculum. The students' desks were arranged in groups and they spent a large part of the class working in their groups. The class started with a set of five problems that involved factoring and expanding polynomials, a highly symbolic type of problem. For the first 30 minutes of class the students worked on these opening problems in groups and then five students presented their solution to one of the problems to the class on a whiteboard. For the remainder of the class period, the topic shifted to probability and students discussed more open-ended problems from the curriculum in groups. Occasionally groups would present their solutions to the problems.

Richard captured a total of 10 clips in which he saw his students thinking algebraically. All but three of the clips were captured during the first 30 minutes of class when the class was working on the symbolic algebra problems. Richard's ideas about algebra and the symbolic nature of algebra prompted him to capture most of the clips where students were working through problems with symbols and manipulating expressions or equations using algebraic structures. For example, he captured one clip while listening to a groups' presentation of their solution to one of the symbolic problems he posed at the beginning of class. However, when asked why he thought he saw his students thinking algebraically in the clips, he discussed nine of the ten clips in terms of students' sense-making and only one of them was discussed from a procedural standpoint. For example, Richard captured the clip of students symbolically solving a problem, yet when he discussed the clip further, he said that the student was thinking algebraically, not because of symbol use, but because of the students' "high-level understanding of exponents". Richard's focus went beyond mere knowledge of the rules to a concern with students' understanding and sense-making. When he discussed the clip, he did not mention the symbol manipulation or procedures but focused on the depth of the student's understanding.

I found it interesting that most of the clips were captured during the very symbolic algebraic part of the lesson yet he discussed most of them from a conceptual standpoint. I would have thought that if he limited his attention to the traditional algebra problems, he might focus on the symbolic or procedural nature of the students' thinking, as Dana and Maggie did. Why did Richard discuss procedural clips differently than the other two teachers? What lead to his focus on the conceptual nature of his students' thinking? I turn to Richard's pre-interview to answer some of these questions.

The theme that emerged from Richard's pre-interview was the dual nature of his goals for students. Because his goals were influenced by both his pre-teaching experiences and his curriculum, he had two distinct goals for his students. He described one of his goals for student learning as helping students master as much of the content as they could. This idea seemed to drive many of the clips he captured. This idea of mastery of the concepts seems to me to be in line with a more procedural idea of algebra and algebraic thinking.

Richard's learning goals for students seemed to be influenced equally by his ideas about algebra and the reform curriculum he uses. When I asked him about his goals for students, he answered, "I think that I'm probably a product of this curriculum. I'm not saying I know what they were thinking but I know my opinion is highly influenced by what I taught over the past five years." His curriculum emphasizes the conceptual nature of student thinking. The dual nature of Richard's goals for students becomes apparent when he talks about his experiences.

Richard's ideas about algebra and the curriculum influenced each other, however they were also in conflict with each other. There was kind of a trade-off, back-and-forth kind of interaction between the two and how they would affect his learning goals for students. His preconceived ideas about algebra affected the way he viewed the curriculum. He appreciated the conceptual nature of the curriculum but also believed it needed to be supplemented with more procedural material. On the other hand, the curriculum affected his ideas about algebra. The dual nature of his influences was apparent again when he said, "I also think that, however…that (the curriculum) is much more on the 'it's really important to conceptualize and understand end of the spectrum' than 'let's do these equations and understand'. You know? So I feel like I'm kind of trying to bridge the two." So his ideas about algebra before he started teaching influenced how he worked with the curriculum and the curriculum he uses influenced his preconceived ideas about algebra. Both factors were influential in his goals for students. This dichotomy might explain the fact that even though Richard tended to record clips in which students were involved in procedural activities, he focused on the deeper, conceptual understanding the students exhibited.

Richard's preconceived ideas about algebra were focused on procedures and skills and he believed that the probability lesson might not provide many instances of this type of thinking. The fact that all but one of the clips he captured was from the part of class where students were engaged in manipulating symbols might illustrate his need for symbols and equations as *signifiers* of algebraic thinking. On the other hand, influenced by curriculum he teaches, when he discussed the clips he captured, he did so in a conceptual way, meaning that he did not discuss the symbol manipulation. Instead he discussed the deeper, conceptual thinking his students were engaged in. Dana also captured clips where students were engaged in manipulating symbols, but when she talked about them, she discussed symbol manipulation and following procedures as the primary reason for identifying the students' thinking as algebraic.

## Conclusion

Using a variety of methods allows us to draw a more complete picture of teachers' conceptualizations of algebra. In particular by using a video camera that allows teachers to capture footage in the moments of instruction, I was able to go beyond what is typically available through interviews and classroom video alone. Classroom activity is complex and watching video of an entire class does not necessarily give us insight into what a teacher was attending to while they were teaching. However, allowing a teacher the ability to capture and save video clips in the moments of instruction has the potential to give us this insight. Furthermore, interviewing the teacher about the clips they captured allows us access to the teacher's insight about the moment they captured and why that moment stood out to them.

I also extended prior research by exploring the relationship between teachers' professed and enacted views about algebra in particular. In line with previous research that has noted how difficult it can be to balance procedural and conceptual views I found that teachers continue to struggle with this dilemma.

In this study I showed that teachers can hold more than one idea about algebra and it is difficult to predict which aspects of their conceptualizations they will draw on from moment to moment while teaching. If we continue to explore teachers' ideas using methods described in this study we may be better able to understand their conceptualizations- where teachers' ideas come from and how they drive what a teacher notices in the classroom. This, in turn will help us facilitate teachers' transitions to a more balanced approach to algebra teaching that includes conceptual as well as procedural activity.

## References

Ainley, J., Luntley, M. (2007). The role of attention in expert classroom practice. *Journal of Mathematics Teacher Education, 10*, 3-22.

Battista, M, & Borrow, C. (1998). Using spreadsheets to promote algebraic thinking. *Teaching Children Mathematics*, *4*(8), 470-478.

Carpenter, T.P., Franke, M.L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in the elementary school.* Portsmouth, NH: Heinemann.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12,* 311-329.

Cooney, T. J. (1985). A beginning teacher's view of problem solving. *Journal for Research in Mathematics Education, 16*, 324 - 336.

Doerr, H.M. (2004). Teachers' knowledge and the teaching of algebra. In K. Stacey, H. Chick, & M. Kendal (Eds.), *The future of the teaching and learning of algebra, the 12th ICMI study* (pp. 267-290). Boston: Kluwer Academic Publishers.

Kaput, J. (1998). Transforming algebra from an engine of inequity to an engine of mathematical power by "algebrafing" the K-12 curriculum. In S. Fennel (Ed.), *The nature and role of algebra in the K-14 curriculum: Proceedings of a national symposium* (pp. 25-26). Washington, DC: National Research Council, National Academy Press.

Kaput, J.J. (2008). What is algebra? What is algebraic reasoning? In J.J. Kaput, D.W. Carraher, & M.L. Blanton (Eds.), *Algebra in the early grades*. Mahwah, NJ: Erlbaum.

Kieran, C. (1992). The learning and teaching of school algebra. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390-419). New York: Macmillan.

Kieran, C. (1996). The changing face of school algebra. In C. Alsina, J. Alvarez, B. Hodgson, C. Laborde, & A. Perez (Eds.), *8th International Congress on Mathematical Education: Selected lectures* (pp. 271-290). Sevilla, Spain: S.A.E.M. Thales.

Kieran, C. (2004). The core of algebra: Reflections on its main activities. In K. Stacey, H. Chick, & M. Kendal (Eds.), *The future of the teaching and learning of algebra, the 12th ICMI study* (pp. 21-33). Boston: Kluwer Academic Publishers.

Lampert, M., & Ball, D.L. (1998). Teaching, multimedia and mathematics: Investigations of real practice. *Journal of Mathematics Teacher Education, 2*(3), 311-319.

Menzel, B., & Clarke, D. (1998). Teachers interpreting algebra; Teachers' views about the nature of algebra. In C. Kanes, M. Goos, & E. Warren (Eds.), *Teaching mathematics in new times* (Proceedings of the 21st annual conference of the Mathematics Education Research Group of Australasia, pp. 365-372). Gold Coast, Australia: MERGA.

Nathan, M.J., & Koedinger, K.R. (2000a). Teacher's and researcher's beliefs about the development of algebraic reasoning. *Journal for Research in Mathematics Education, 31*(2), 168-190.

Nathan, M.J., & Koedinger, K.R. (2000b). An investigation of teachers' beliefs of students' algebra development. *Cognition and Instruction, 18*(2), 209-237.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Reich, S., Goldberg, L., & Hudek, S. (October, 2004). *Deja View Camewear Model 100*. Paper presented at the First ACM workshop on Continuous Archival and Retrieval of Personal Experiences. New York.

Rosean, C.L., Lundeberg, M., Cooper, M., Fritzen, A., & Terpstra, M. (2008). Noticing noticing. *Journal of Teacher Education, 59*(4), 347-360.

Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction*. *20(2), 119-150.*

Sherin, M.G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.) *Video research in the learning sciences* (pp. 383-395). Hillsdale, NJ: Erlbaum.

Sherin, M. G., Russ, R. S., Colestock, A. A. (in preparation). Accessing mathematics teachers' in-the-moment noticing. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.) *Mathematics teacher noticing: Seeing through teachers' eyes.* New York: Routledge.

Sherin, M. G., Russ, R., Sherin, B., & Colestock, A. (2008). Professional vision in action: An exploratory study. *Issues in Teacher Education*.

Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education, 13*(3), 475-491.

Stephens, A. (2008). What "counts" as algebra in the eyes of preservice elementary teachers? *Journal of Mathematical Behavior, 27*, 33-47.

Usiskn, Z. (1988). Conceptions of school algebra and uses of variables. In A. Coxford (Ed.), *The ideas of algebra, K-12* (1988 yearbook, pp. 8-19). Reston, VA: National Council of Techers of Mathematics.

# Conceptual Change and Epistemic Growth through Reflective Assessment in Computer-Supported Knowledge Building

Carol K.K. Chan, Ivan C.K. Lam, University of Hong Kong, Pokfulam, Hong Kong
Email: ckkchan@hku.hk; ivanlmhk@gmail.com

**Abstract:** This study examined the design and process of how students' reflective assessment promoted collaborative metacognition for conceptual and epistemic changes, mediated by Knowledge Forum, a computer-supported environment. The design involved augmenting knowledge-building inquiry with reflective assessment – Students wrote reflective summaries to track their initial understanding and trajectories of growth toward scientific understanding in the domain of electrochemistry. Two classes of 10th grade students in Hong Kong participated in Reflective Assessment (RA) and Reflective Assessment with Scaffolds (RAS) conditions. Results indicated that both classes improved on conceptual-change and epistemic-beliefs measures; the effects were stronger for the class with scaffolds. Qualitative analyses showed how students' reflective assessment and collaboration helped them to develop metaconceptual and epistemic awareness as they examined their own and others' understanding. A path analysis indicated that students' engagement on Knowledge Forum predicted reflective collaboration that in turn exerted effects on their changes in conceptual understanding and epistemic beliefs.

## Theoretical Perspectives

Across science education and learning sciences, there are now major shifts towards theories of learning that emphasize both individual and social aspects of science learning (Duit & Treagust, 2003; Vosniadou, 2008). It is now widely accepted that science learning can be facilitated when students articulate their prior ideas and explain their understanding to each other. Conceptual change is examined emphasizing the social construction of knowledge and discursive interactions in classroom (Scott, Asoko & Leach, 2007). Furthermore, researchers now question conceptual change as a sudden change or a replacement of misconceptions with scientific ones through externally-driven conceptual conflict (Chan, Burtis & Bereiter, 1997). Instead, conceptual change involves a gradual process where science concepts are gradually restructured mediated by students' intentional-learning strategies (Sinatra & Pintrich, 2003). From a learning sciences perspective, there is a need to capture processes of conceptual change supported by collaborative discourse and to design learning environments to foster intentional conceptual change.

### Intentional Conceptual Change

Current research on intentional conceptual change emphasizes the role of learners' metacognitive strategies, epistemic beliefs and agency in knowledge restructuring (Sinatra & Pintrich, 2003). It also points to the need to designing learning environments that encourage learners to employ goal-directed, reflective strategies and to develop metaconceptual awareness. Cognitive research has shown that students' epistemic beliefs can constrain or facilitate student thinking, reasoning, and science learning (Stathopoulou, & Vosniadou, 2007). Researchers have argued that conceptual change involves not only changes in concepts; there needs to be changes in students' views about the nature of science (Duit & Treagust, 2003). Vosniadou (2008) noted that conceptual change involves metaconceptual awareness – Students will be able to learn science concepts and principles only if they are aware of their prior understanding and the shift of their initial views toward science knowledge. Increasingly the emphasis is to examine conceptual change that includes not only individual cognitive development but also social and collective aspects; socio-cognitive discourse plays a key role in facilitating conceptual change. Although there has been much progress indicating the role of metacognition and epistemic beliefs on students' conceptual change, most of the research are correlation studies. Fewer studies have examined designing for intentional conceptual change that brings about metaconceptual awareness with epistemic changes supported by social and collective discourse.

### Computer-Supported Collaborative Knowledge Building

How can current views of intentional conceptual change be integrated in instruction emphasizing collaborative discourse? Researchers have suggested how computer-supported collaborative learning (CSCL) can make important contribution to conceptual-change research (Miyake, 2007). Relating to the roles of interaction and discourse in science classroom, technological affordances can further provide a medium whereby students can articulate, communicate, represent and reconstruct their ideas for sustained inquiry. This study adopts an educational approach, knowledge building, a forerunner of CSCL, that emphasizes knowledge creation as a collective work of the community; and that knowledge is improvable by means of a discourse (Scardamalia & Bereiter, 2006). To support student discourse, knowledge Forum (KF), a multi-media database constructed by students, was designed to support collaborative knowledge building discourse. In a knowledge-building community (both face-to-face and online), students engage in scientific discourse that involves posing cutting-edge problems, generating theories and conjectures, searching for scientific information, elaborating on others' ideas, co-constructing explanations, and revising their theories. Students' pre-instructional ideas and learning pathways can be represented on the computer forum and thus become objects of inquiry for conceptual change. There is now substantial evidence on the roles of knowledge building on students' collective inquiry and scientific understanding (e.g., Zhang et al., 2007).

### Knowledge Building, Reflective Assessment and Conceptual Change

A question may then arise as to why another study is needed on knowledge building and scientific understanding. Despite major progress in two decades of research, there have been no systematic studies using this knowledge-building approach to examine conceptual change. Various principles advocated by researchers in conceptual change such as intentional goal-directed strategies, metaconceptual awareness, epistemic beliefs (Vosniadou, 2008) are well aligned with knowledge-building. However, how collaborative knowledge-building dynamics can bring about metaconceptual awareness and epistemic changes has not been examined. Further, we argue that knowledge building can enrich conceptual change studies that often emphasize small-group collaboration. There is a need to understand how conceptual change can take place in communities of learners and knowledge-builders. Knowledge building is not just a pedagogical approach but a theory of epistemology; so how students working with knowledge might change their epistemic views are fruitful lines of inquiry. Finally, knowledge-building research on science learning has mostly been conducted with elementary-school children. It would be useful to extend the scope of inquiry to investigating knowledge building for high-school science.

This study employed a design developed in research on assessment of knowledge building with students assessing their own collaboration (Lee, Chan, & van Aalst, 2006; van Aalst & Chan, 2007). Research has shown that students assessing their own scientific inquiry promoted metacognition (White & Fredericksen, 1998). Similarly, student-directed e-portfolio assessment with students documenting how they collaborated in knowledge-building discourse fostered their domain understanding (van Aalst & Chan, 2007). This study extends this line of inquiry: We designed knowledge building for conceptual change focusing on promoting metacognition in a collaborative context. Extending our earlier study (xx), we asked students to reflect on their prior conceptions and to track how they moved towards scientific understanding as they considered others' contributions and revised their ideas.

To iterate, this study aimed to design and examine how reflective assessment with collaborative dynamics would promote metaconceptual and epistemic awareness for conceptual change. Three research questions were included: (1) What was the role of the knowledge-building environment on students' conceptual and epistemic changes? (2) How did students' reflection contribute to their changes in conceptual and epistemic understanding? and (3) what were the relations among knowledge-building dynamics, conceptual change and epistemic growth?

## Methods

### Participants

Eighty 10th graders (Age ranging 15-16) in two chemistry classes in Hong Kong participated in this study. The lessons were conducted in English and students wrote notes in English on Knowledge Forum (KF). Both classes engaged in knowledge-building inquiry with reflective assessment. The first class is called Reflective Assessment (RA, n = 40) and the second Reflective Assessment with Scaffolds (RAS, n = 40). Both classes were taught by the same teacher, who had taught high-school chemistry for more than twelve years and had used knowledge-building pedagogy for over 6 years.

### Procedure

This study was conducted in the second semester of 2008-09 lasting from Feb-June (16-18 weeks). There were five chemistry lessons each week; each lesson was of 35 minute duration. In both classes, students learned electrochemistry using knowledge-building inquiry approaches and they wrote computer notes on Knowledge Forum. Due to school policy for comparable curriculum, students had similar inquiry experiences (see later section, 1-4) and both used the same instructional topics, textbooks and reference materials and conducting the same experiments. However, there were also some key differences: Primarily students in the "Reflective Assessment with Scaffolds" class (RAS) wrote reflective summaries on KF using a set of pre-designed conceptual-change "scaffolds" (e.g., *My initial ideas, What we think, What I think now).* Alternatively, students in the "Reflective Assessment" class (RA) wrote reflective summaries without the scaffolds (RA). "Scaffolds" are prompts or sentence openers on

KF; different scaffolds can be designed depending on subjects and contexts. The use of conceptual-change scaffolds were to examine further whether the use of scaffolds would enhance further students' collaborative reflection.

## Designing a Knowledge-Building Environment for Conceptual Change

We designed the learning environment based on knowledge building pedagogy aligning that with conceptual change principles (Vosniadou & Kollias, 2003). The principles and activities of the design are briefly described:

(1) *Activate prior knowledge through classroom discourse.* Students need to activate and reflect on prior knowledge and to articulate their ideas for science learning. Students worked in groups discussing science topics/problems in classroom. Students were scaffolded to present their ideas, make observations of experiments; raise questions they did not understand, elaborate and comment on others' views. Students' ideas were shared and made public using concept-maps, posters and knowledge-building walls (boards for posting ideas). When students became familiar with articulating their thinking, they then continued their inquiry and contributed their ideas and questions onto KF.

(2) *Foster metacognition through KF scaffolds and problem-centered inquiry.* Students were encouraged to raise authentic problems from prior ideas and daily life on electrochemistry that puzzled them. Extending the classroom inquiry, students wrote their ideas on the discussion views on KF (Figure 1 left): Students engaged in goal-directed inquiry- they posed problems, made conjectures/hypotheses, co-constructed explanations, compared different explanations and revised their understanding. KF scaffolds including '*I need to understand*', '*my theory'*, '*new information*', and '*a better theory'* prompted metacognitive thinking and theory revision. Teacher facilitation involved helping students integrate KF and classroom discourse and to notice conflicts, discrepancies and identifying gaps for further inquiry.

(3) *Develop deep understanding through model-based explanatory inquiry.* Students were involved in constructing models of chemical cells using everyday materials and explaining to each other how their chemical cells work.

(4) *Integrate fragmented ideas and use 'rise-above' explanation.* To tackle the problem of fragmented ideas, a common barrier to conceptual change, students deepened their understanding using KF functions of 'rise above' and 'references' to synthesize diverse and fragmented ideas from classmates as they worked towards more *coherent explanations.* Teachers integrated online discussion with classroom talk to help students deepen their inquiry.

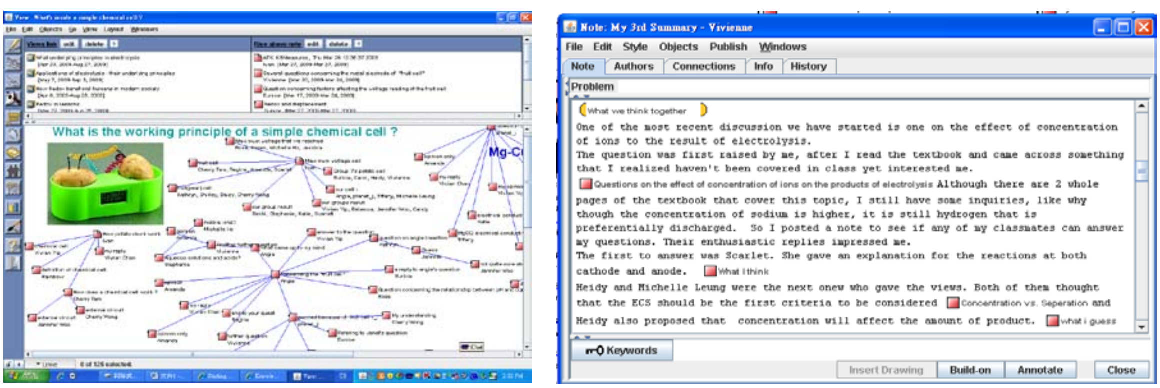



Figure 1. A Discussion View on Knowledge Forum (left) and an Example of a Reflective Summary Note (right).

(5) *Develop metaconceptual awareness through reflective assessment.* In both classes, students were asked to review the notes on KF, to reflect on their initial beliefs, and to track their changing ideas incorporating classmates' ideas (Figure 1, right): The teacher prompt written on KF included: You are encouraged to review the computer notes written by you and your classmates in the database. Write a summary note to reflect and to consolidate what you have learnt from the views of '*Batteries' &* '*Simple chemical cell*'…In writing the summary, you may select relevant computer notes (references) as evidence that support your understanding. Think about how your chemical knowledge has developed or changed. As noted above, in the Reflective Assessment with Scaffolds (RAS) class, students were further provided with conceptual-change scaffolds (e.g., "My initial idea"). Students could also use other KF scaffolds such as "a better theory" and "this theory cannot explain" and "putting our knowledge together".

To help students with their reflection, in both classroom and online discourse, the teacher would scaffold students to reflect more deeply in discussing their notes. As well, the teacher also wrote on KF making some observation: "So far I have observed some benefits of chemistry gains from your summaries: (i) You are encouraged to give more evidence (chemical knowledge) to reflect on your understanding or knowledge gains (not just say KF is good). (ii). It may be useful to tell us more of your initial thoughts and final thoughts on a particular chemistry concept. (iii) Continue to build on others' notes to deepen your ideas after reviewing these previous notes. Of course you can continue to raise other (emerging) questions or observations in everyday life."

## Analyses and Results

We first assessed student changes in conceptual and epistemic changes after the instruction; we then examined why and how such changes might take place relating to knowledge-building inquiry and reflection, and finally we investigated the relations among knowledge building activities, conceptual-change learning and epistemic changes.

### Conceptual-Change Learning in Electrochemistry

Conceptual-Change Written Tests and Explanation Written questions were designed to tap into students' understanding of electrochemistry probing into students' alternative conceptions. The written questions were designed based on studies on conceptual-change in electrochemistry in science education research. Questions included open-ended ones asking students to explain their understanding of key concepts to probe their conceptions (what is oxidation?). Responses to forced-choice questions were coded on students' explanations of their choice. The pretest and posttest percentage scores were 13.8 (7.1) and 57.3 (12.3), and 14.1 (6.4) and 49.1 (15.4) for Reflective Assessment with Scaffolds (RAS) and Reflective Assessment (RA) classes, respectively. Analyses of pre-posttests using paired t-tests showed that both classes improved on the conceptual-change tests with RAS, $t$ (39), = 20.14, $p$<.001, and RA, $t$ (39), = 14.21, $p$<.001. Analyses using ANCOVA controlling for pre-test differences indicated that higher scores were obtained for students in Reflective Assessment with Scaffolds (RAS).

Changes in Alternative Conceptions in Written Tests The number of alternative conceptions (misconceptions) for each student was identified from their responses to the written tests. The means of alternative conceptions were 6.4 (1.7) and 5.7(1.2) at pretests and 1.9 (1.5) and 3.0 (1.9) at posttests for RAS and RA classes respectively. Students in both classes decreased in the number of alternative conceptions after instruction. ANCOVA on posttest results using pretest scores as covariates indicated that there were more decreases in the number of alternative conceptions for students in the class of Reflective Assessment with scaffolds (RAS), $F$ (1, 78) = 10.08, $p$<.003.

### Changes in Epistemic Beliefs

Students were administered a questionnaire on epistemic beliefs (adapted from Conley et al., 2004) to examine the ways students think about the nature of knowledge. The questionnaire was pilot tested with over 300 students and validated in our earlier study (xx). Consistent with the earlier study, three factors were identified (1) "Certainty-Source", (2) "Development" and (3) "Justification" with scale reliabilities ranging from 0.64 to 0.83. An example of an item on Certainty is "Most questions in science have one right answer"; an example of an item on Development is "Scientific knowledge will not change over time", and an example of an item on Justification is "Ideas in science can come from your own question". Interviews were also conducted to provide more information on what students thought about these items; qualitative analyses are currently undertaken. Paired t-tests showed that students in both classes made changes towards more sophisticated epistemic beliefs based on the combined scores, for RAS, $t$ (38) = 4.73, $p$<.001 and for RA, $t$ (38) =2.27, $p$<.03. Separate analyses showed that students in RAS improved more on the subscales of "certainty-source" and "development" and students in RA improved more on "justification".

### Students' Knowledge Forum Engagement and Collaborative Reflection

Knowledge Forum Engagement (ATK) We examined students' overall participation and engagement in Knowledge Forum (KF) using a software called the Analytic Toolkit (ATK, Burtis, 1998) that uses log files to show students' participation and activity on KF. We included several ATK indices commonly used in knowledge-building research : (1) number of KF note written, (2) % of notes read, (3) % of linked notes, (4) scaffolds (thinking prompts) (5) keywords and (6) revision. Some of these indices show student collaboration (e.g., notes read/linked) and others reflect metacognition such as the use of scaffolds" and "revision" of notes for purposeful activities.

We now report results of ATK indices for RAS and RA classes: number of notes written, 21.5 and 22.60 notes; percentage of notes read, 55% & 47%; note-linked, 80% and 68%; and scaffolds, 21 and 9.0, respectively. Although there were no norms for ATK, comparison with other studies indicated that these students were engaged actively participating and collaborating on KF. Compared to the literature on online learning with fragmented contribution (Hewitt, 2003), these number indicate high level of contribution; they are also comparable to those identified in mature knowledge building communities (Zhang et al., 2007). Comparison of ATK indices also showed some advantages for the RAS class over the RA class.

Reflective Assessment and Collaborative Reflection As described above, students wrote three reflective summaries to reflect on their initial and new ideas in electrochemistry based on their discourse. These reflective summaries were scored on a 6-point scale. At the lower levels (1-2), the reflection depicts that students were just describing new information with limited reflection on what it meant for their conceptions. At the mid-levels (3-4), students demonstrated some personal thinking for identifying misconceptions or knowledge gaps in their understanding. At the highest levels (5-6), student reflected on their prior knowledge, identified gaps, considered how others' ideas supported their reflection; they demonstrated metaconceptual awareness of initial and new ideas and noted how they changed in their understanding of some concepts The summaries were coded and currently inter-rater reliability checks were being conducted. In the following, two contrastive examples of reflective summaries are provided to suggest how collaborative reflection may foster metaconceptual awareness for conceptual and epistemic growth..

Table 1: An example of a reflective summary note (low-level response).

| Restating Information | Excerpts from the Reflective Summary Note |
|---|---|
| States impartial information | In the simple chemical cell, I found out that a potato cell can actually conduct electricity and drive the calculator to work. … |
| Makes reference to one note; no reference to one's own thinking | Through Florence's note, I knew that the electromotive force within each potato makes to move electric current. And the copper wire makes the electrons move in the potato, causing energy to move into the clock. This let me know more about how a potato cell conducts electricity |
| Describes factual information and formulae; no reflection | In the redox reaction of copper, I knew that …when the copper reacts with oxygen, copper acts as the reducing agent and causes oxidation, while oxygen acts as the oxidizing agent, causes reduction : $2Cu(s) + O_2\ (g) \rightarrow 2CuO\ (s)$<br>And when hydrogen reacts with the copper oxide, hydrogen acts as the reducing agent, reduce the copper oxide : $CuO(s) + H_2(g) \rightarrow Cu(s) + H_2O(l)$… |

Table 2: An example of a reflective summary note (high-level response).

| Reflective Metacognition | Excerpts from the Reflective Summary Note |
|---|---|
| The student reflected on her understanding on state of matter and electrolysis | |
| - Identifies her prior conception; reflects on the source of difficulty<br>-Considers role of textbook information | What I think before The products of electrolysis are always the same as long as the chemical salt is the same, disregarding its state.<br>-*Because in many previous textbook chapters such as molarity, water plays no role* in the sense that it does not react.[1] What I think I thought we do not melt the salt in electrolysis just because it is too troublesome. |
| -Considers other ideas and explanations<br>-Selects relevant ideas and organizes them to show some learning pathways (e.g., further) | What we think together Jennifer provides detailed chemical equations to explain the difference[2] Agree. Candy *further* provides the significance of the difference[3] *molten and aqueous*, that metals like sodium could never be formed in electrolysis if there were no molten sodium salts…She also mentioned an interesting fact that mercury electrode can be used to extract pure sodium. This thought is *further worked* on.[4] Mercury electrode |
| -Reflects on new idea<br>–Continues to query gap of understanding | When doing work regarding electrolysis, I have to look carefully whether the chemical is *molten or aqueous* as the results are very different. *However I still do not understand* the working principles of mercury electrodes. |
| The student noted another cycle examining factors influencing electrolysis | |
| -Reflects on prior beliefs; notes prior gaps in understanding;<br>-Identifies source of confusion & difficulties<br>-Refers to textbook as conveying information not in real science | What I think before As stated above, *I thought* the products of electrolysis are always the same if chemical salt is the same. *I didn't think a* higher voltage, except speeding up the process, will produce other results. For example, in aluminum anodization, *I thought only* the quality of the original aluminum will provide different results. Moreover, although I noticed that the metal deposited on the electrode is unevenly distributed, *I always thought* it was due to our poor skills or equipment. I have not considered it a natural occurrence*, mostly because textbooks often show the electrode fully and smoothly covered with the metal.* |
| -Considers others' ideas<br>–Identifies puzzling info | What we think together …Rainbow suggested that temperature as well as acidity affects the results. *She also gave a curious suggestion that lower temperature gives* |

| | |
|---|---|
| and reflects on what she does not understand;<br>- Examine various ideas and explanations to help her move toward better Understanding | *thicker layer of aluminum oxide, which I still can't understand* as I thought a higher temperature facilitates reaction, like what angie said [6] Answer.. [7] more information Cherry Wang further told us that an unsmooth layer is resulted as the metal plated is attracted to external corners and avoids internal ones. [8] Effects and Limitations of electrolysis. Besides, a higher current can lead to the formation of other substances in the solution, as mentioned in [9] 2 substances formed? |
| Reflects on new understanding | What I think now The product of electrolysis could be affected by various [external] effects; the metal plated does not naturally spread out evenly. |
| The student included another scaffold and reflect on her beliefs and knowledge | |
| Reflects on beliefs about knowledge pointing to coherence (structure of knowledge) and evidence in justification | My belief on learning &knowledge The most useful part is it broadens our thinking by *relating one topic to another*. For example I have never considered rusting from a redox or electrolysis point of view. *I believe we could learn more if we try to search for information before writing the notes instead of guessing without any evidence* |

Note: 1. What I think – Scaffolds in reflective summaries
2. [3] Reference notes in summaries with hyper-links to students' notes in the database

Table 1 shows an example of a reflective summary in which Student A was not actually engaged in reflection – She described some impartial information; referred to only one note from a classmate, described some factual information and formulae but did not reflect on her understanding or made attempts to describe changes. Table 2 shows another example with two related episodes – In the first one, Student B identified her prior understanding (state of matter & electrolysis); it is a key concept and a common alternative conception. She reflected on why she had the problem (prior knowledge & textbook); considered various classmates' explanations and organized them; and she noted her new understanding but continued to identify areas she did not understand (mercury electrode). In the second episode (factors influencing electrolysis), Student B continued to identify her prior ideas and gaps of understanding; noted others' ideas and she puzzled over her classmate's curious information. She tracked different ideas but focused on the original problem and reflected on her new understanding. Student B employed good use of the scaffolds and demonstrated metacognition noting what she knew and what she did not understand. As well, she showed metaconceptual awareness as she became more aware of her initial conceptions and how they differed from scientific ideas. There are various instances that showed how such reflection prompted Student B to examine nature and source of knowledge. For example, she noted textbook as unauthentic science and "imperfect" source of knowledge. Furthermore, she concluded using a scaffold My belief on learning &knowledge implying some thinking about the importance of coherence (structure of knowledge) and role of evidence.

### Contribution of Knowledge Building Activity to Conceptual & Epistemic Changes

Correlation Among Measures Table 3 shows the correlation among various measures; we pooled the two classes because the key principles of reflective assessment are same. To improve the coherence for analyses, the various ATK KF indices were combined using factor analysis (see Lee et al., 2006). Two factors were extracted, Factor 1 is called **metacognition** (scaffold use, note revision and keyword) that explains 32.6% of the variance, and factor 2 is called **collaboration** (notes created, notes linked and read) that explains 30.1% of the variance. Table 3 indicates that KB reflection was significantly correlated with ATK metacognition, conceptual change and epistemic change scores. Furthermore, ATK collaboration scores were significantly correlated with conceptual-change scores.

Table 3: Correlations among ATK forum participation, KB reflection, conceptual and epistemic measures

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1.ATK Metacognition | | | | | | |
| 2 ATK Collaboration | | | | | | |
| 3. KB Reflection | .57*** | .18 | | | | |
| 4. Post-Epistemic Belief | | | | | | |
| 5. Epistemic Belief Change | | | .27* | .26* | | |
| 6. Post-Conceptual | | .38** | .44** | | | |
| 7. Conceptual-Test Change | | .31* | .40** | | | .90*** |

Note: * $p<.05$; ** $p<.01$; *** $p<.001$

Regression and Path Analyses We conducted hierarchical regression analyses on students' post-conceptual scores first using exam results and pre-conceptual scores (prior achievement) as predictors ($R^2$ =.34) When KF engagement (ATK Collaboration) was added, additional 4.6% variance was explained ($R^2$ =.38). Further, when we added reflection scores (summaries), there were additional 7.4%variances explained ($R^2$ =.46); R2 changes were all significant. These results indicated that over and above science achievement and prior knowledge, student engagement in forum and metacognitive reflection further contributed to post-test conceptual scores (Table 4)..

Table 4: Regression on post-conceptual scores with achievement, ATK collaboration & reflection as predictors.

| | R | $R^2$ | $R^2$ Change | F Change |
|---|---|---|---|---|
| Mid-year exam & pre-conceptual scores | .58 | .338 | .338 | $39.82^{***}$ |
| Forum – ATK collaboration | .62 | .384 | .046 | $5.63^{*}$ |
| Collaborative Reflection | .68 | .458 | .074 | $10.3^{**}$ |

Note: *p<.05; **p<.01; ***p<.001.

We also employed a path analysis testing a causal model to provide a more coherent picture: Student engagement in Knowledge Forum including both ATK metacognition (e.g., scaffold use; notes revision) and ATK collaboration (e.g., notes read and linked) predicted KB collaborative reflection that further exerted effects on students' epistemic belief change scores and conceptual-change scores.

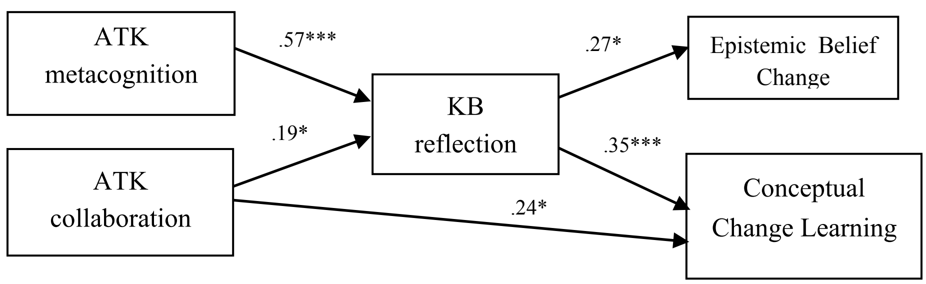



Figure 2. A Path Analysis Indicating Contributions of Knowledge-Building Activity to Conceptual and Epistemic Changes.

## Conclusions, Implications and Significance

This study developed a knowledge-building environment augmented with reflective assessment to examine and to foster conceptual change and epistemic growth. The environment was designed in ways that aligned with conceptual-change principles focusing on students' goal-directed strategies and agency elaborated with collaborative reflection. Results indicated that after the instruction, students in both classes changed towards more sophisticated scientific understanding measured by conceptual-change tests in electrochemistry. There is some evidence suggesting students also made some shifts in their epistemic beliefs. Comparing the two classes when students both had reflection, the effects were stronger for students supported with conceptual-change scaffolds for reflection.

A key question to address is to examine knowledge-building dynamics and to explain the pre-posttest gains, that is, how it may be possible for students to experience these changes. We demonstrated that asking students to engage in knowledge-building inquiry augmented with assessing their own scientific understanding in their discourse could help them activate prior knowledge, develop meta-conceptual awareness, and reconstruct fragmented views into more coherent explanations. As the excerpt shows, the student identified her prior conceptions; examined the nature of difficulties; considered others' views; puzzled over gaps of understanding; and integrated fragmented ideas into a more coherent account. The excerpt also suggests metacognitive reflection was facilitated because of the rich collaborative context with diverse ideas from classmates. Although ATK KF indices are quantitative, the extent to which students engaged in KF was a prerequisite for conceptual work – It supports the notion of how CSCL may provide the medium through which students can articulate, represent, interact and inquire into their ideas for sustained inquiry. Regression analyses provided support of this account indicating that students' engagement on KF and collective reflection predicted post-test conceptual scores *over and above* their academic achievement. A path analysis further showed that students' activity on Knowledge Forum predicted their collaborative reflection that in turn influenced students' posttest conceptual and epistemic changes.

Regarding epistemic shifts from questionnaire data, the excerpts provided some glimpses suggesting how changes were possible in collaborative inquiry and reflection. For example, Student B reflected upon the textbook as a source of 'imperfect" knowledge decontexualized from real-world science (i.e., textbook pictures always show smooth surface). She also pondered upon the importance of the connection among topics (structure of knowledge) and the need for justification using evidence. When students tackled authentic problems and reflected on their changes in understanding, they might be better able to see that knowledge is not certain and that it can be advanced.

To conclude, this study integrated two lines of research on intentional conceptual change and knowledge building examining how conceptual and epistemic changes could be fostered through collective inquiry and reflection. The study furthers research on knowledge building aligning with conceptual change augmented with student-directed assessment. Whereas many studies have demonstrated the role of metaconceptual awareness and epistemic beliefs, this is one of the few studies that illustrated how it might be possible to foster conceptual, metaconceptual and epistemic changes through collective reflection. Furthermore, the conceptual-change reflective assessment summaries provided a rich data source to examine students' alternative understanding of electrochemistry, as well as to track how students could move between individual and collective growth towards conceptual change. Instead of eradicating misconceptions, students' prior understandings could become objects of collective inquiry. Metacognition for conceptual change is not an individual accomplishment but one that can be examined and fostered in a community of knowledge builders. Turning over agency to students and asking them to reflect and to assess their own and collective understanding may help them develop conceptual and epistemic changes and provide insights into further research on collaborative conceptual change.

## References

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Lawrence Erlbaum Associates.

Burtis, P.J. (1996). Analytic Toolkit for Knowledge Forum. Ontario Institute for Studies in Education/U of Toronto.

Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction*, *15*, 1-40.

Conley, A.M., Pintrich, P.R., Vekiri, I. and Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology*, *29*, 186-204.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*, 671-688.

Hofer, B.K. and Pintrich, P.R. (Eds.) (2002), *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lee, E.Y.C., Chan, C.K.K. and van Aalst, J. (2006). Students assessing their known collaborative knowledge building. *International Journal of Computer-Supported Collaborative Learning*, 1, 57-87

Miyake, N. (2008). Conceptual change through collaboration. In S. Vosniadou (Ed.). *International handbook of reesarch on conceptual change* (pp 453-478). New York: Routledge.

Sinatra, G.M. and Pintrich, P.R. (Eds.). (2003). *Intentional Conceptual Change*. Mahwah, NJ: LEA.

Scardamalia, M. and Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of Learning Sciences*, 3(3), 265-283.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97-115). NY: Cambridge University Press.

Scott, P., Asoko, H. and Leach, J. (2007). Student conceptions and conceptual learning in science. In S.K. Abell and N.G. Lederman (Eds.) (pp. 31-56). *Handbook of research on science education.*

Stathopoulou, C. and Vosniadou, S. (2007). Exploring the relationship between physics-related epistemological beliefs and physics understanding. *Contemporary Educational Psychology*, **32**, 255-281.

van Aalst, J., & Chan, C. K. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *The Journal of the Learning Sciences, 16*, 175-220.

Vosniadou, S. (Ed.). (2008). *International handbook of research on conceptual change.* New York: Routledge.

Vosniadou, S. & V. Kollias (2003). Using collaborative, computer-supported model building to promote conceptual change in science. In E. De Corte, L. Verschaffel, N. Entwistle & J. Van Merrienboer (eds.), (pp. 181- 196), *Powerful learning environments: Unravelling basic components and dimensions*. Amsterdam: Pergamon.

White, B. & Frederiksen, J. R (1998). Inquiry, modeling and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*. 3-118.

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9- and 10-year olds. *Educational Technology Research and Development, 55*, 117-145.

# Talking with your mouth full: The role of a mediating tool in shaping collective positioning

Kate T. Anderson, National Institute of Education, 1 Nanyang Walk, Singapore 637616, katherine.anderson@nie.edu.sg
Melissa S. Gresalfi, Indiana University, 1900 E. 10th St., Eigenmann 532, Bloomington, IN 47405, mgresalf@indiana.edu

**Abstract:** Research on classroom discourse has long underscored the integral role of social interaction in learning, yet designing for and supporting meaningful classroom interaction often remains a challenge. This paper examines the ways that elementary school students are positioned and position themselves during activities aimed at promoting reflexivity about small-group math discussions. We examine one student group's participation in three such activities over eight months' time and specifically highlight the role of a mediating tool that was designed to support participation-centered rather than ability-centered conceptions of learning. Our aim is to examine how one focal group's math discussions were or were not shaped by the tool in ways that were consequential for productive collaborative engagement.

## Introduction

The role of small group discussion in shaping mathematics learning has received much attention in the last decade and has been promoted as an important element of effective teaching and learning (Boaler & Staples, 2008; Gresalfi, 2009; Horn, 2005; Staples, 2008). Specifically, authors have pointed to the potential of collaborative interactions in promoting deep conceptual understanding, and more specifically, in supporting the high functioning of heterogeneous groups (Ben-Ari, 1997; Cohen & Lotan, 1995). However, *how* to foster students' learning through small group discussions that require exploring multiple perspectives, negotiating tensions, and probing disagreements continues to be studied; doing so entails changing many of the traditional practices of mathematics classrooms.

This paper considers the affordances and constraints of a particular mediating tool—conversation rubric—that was used to support small group discussions about mathematics. Specifically, we examine how the conversation rubric was taken up by one student group across multiple events in a fifth grade classroom. We analyze their discussions for evidence of whether the mediating tool a) supported collaborative engagement in math activities, and b) shaped students' reflection about their understanding and their positions in discussions in ways that might not have occurred without this particular tool. By systematically examining interactions across a suite of math activities focused on student discussion, we interrogate ways that particular activity designs can afford productive tensions between students that lead to positioning in light of math and each other. The analysis we present constitutes both evidence for how artifacts can shape the texture of learning across math discussions as well as a methodological lens by which to examine math learning discursively across cases and time.

The aim of our analysis is to examine the role of an instructional scaffold as a mediator of collaborative activity. In particular, we consider whether and how students' positioning relative to one another and relative to the subject matter, elementary mathematics, is shaped by externally structured group reflection on recently completed activities. In so doing, this study has both theoretical and pragmatic implications. With respect to theory, we seek to extend positioning theory to consider how tools can serve to mediate interaction. Thus, rather than considering positioning to be an interactive accomplishment of interlocutors alone, we demonstrate how an artifact can serve as a resource and an authority for co-positioning. With respect to practice, we demonstrate how instructional scaffolds structure interaction processes among learners (Kobbe et al., 2007), although not always in intended ways.

## Positioning Theory

Our analysis draws on positioning theory as a means of tracking changes in student collaboration. We consider positioning as moment-by-moment interaction, which involves the kinds of opportunities that interlocutors create for one another to participate in particular ways (Davies & Harré, 1999). This can be seen most easily in talk as participants speak to and about each other (van Langenhove & Harré, 1999). Positioning concerns the fluid arrangements that people make available for themselves and others through their talk and refers to participants' fluctuating status relative to one another. For example, students are positioned relative to each other as smart vs. slow, or cooperative vs. competitive. Positioning can include students' relative academic status but can also reference students' positions outside of the classroom

(Cohen & Lotan, 1995). Interpersonal positioning can also be accomplished though implicit and explicit comparisons between students. Students' ideas might be positioned as equally valuable (even if they are not both accurate, c.f. Lampert, 1990), or one idea might be positioned as less important than another—for example, in a classroom that is more competitive. In some classrooms, distinctions between who is "smart" and who is not may become a particularly dominant form of positioning, while in others such distinctions are rarely made. The effects of positioning can begin to endure as students become associated with specific ways of participating in classroom settings (e.g., Anderson, 2009; Holland, Skinner, Lachicotte, & Cain, 1998; O'Connor & Michaels, 1996; Wortham, 2004).

## The Importance of Math Talk

Significant attention has been given to the challenge of promoting robust mathematical conversation in classrooms. In particular, the goal of such initiatives has been to shift mathematics away from a procedural, didactic activity to an active process that involves strategic thinking, problem solving, decision-making, and talking about math. The challenges of organizing mathematics classrooms in this way have been well documented and include teachers' own histories and experiences with mathematics, their vision of the nature of mathematical thinking and learning, and students' previous histories with mathematics and their resulting expectations for activity in the classroom.

It is clear that classrooms cannot promote problem solving, discussion, and justification without supporting students to learn how to talk with and work with each other. Productive collaboration has been notoriously difficult to facilitate, particularly in classrooms that are organized heterogeneously. However, examples of successful collaboration in such classrooms paint a picture of the possible and a vision of what to strive for (Boaler & Staples, 2008; Staples, 2008). This research has led to recommendations for practice that are likely to support collaboration, including assigning students roles (Ben-Ari, 1997; Cohen & Lotan, 1995), structuring students' interaction (Kobbe et al., 2007), posing complex problems (Staples, 2008), and being consistent about expectations and accountability (Gresalfi, Martin, Hand, & Greeno, 2009). What is clear is that successfully supporting collaboration is no mean feat and requires facility with content and with orchestrating student practice. In this paper, we consider the role of a small intervention that targeted only student group work (as opposed to teacher practice) as a resource for organizing collaborative activity.

Our perspective on the role of discourse in mathematics classrooms aims to make visible the ways that what we say, how we say it, and how we implicate others' engagement in such interactions all shape participation in ways that affect students' opportunities to learn. Supporting productive modes of interaction and awareness of the crucial role of discursive exploration of math topics is central to views of language and learning as inherently social processes (Barwell, 2005). Implications of such sociocultural views on positioning within mathematical learning point to a need to foster and examine discourse that supports learning how to navigate both ways of being (social selves) in relation to ways of knowing (math learners) that are viewed as relevant, expected, or valued in certain classrooms.

In this paper we build on our understanding of student positioning with the goal of investigating how an external artifact mediated student interaction. Considering one group's interactions over the course of the academic year, we track students' patterns of participation over time and examine micro-interactional details of their conversations across three days spanning the ten activity days we conducted throughout the school year. We view such small-group math discourse and the meta-discourse through which students evaluate their participation as evidence of their positioning relative to each other and the math at hand. We thus consider how the tool afforded or constrained interpersonal and perhaps disciplinary positioning.

## Context of the Study

The data we consider draws from the second of a three-year, U.S. National Science Foundation-funded research project that examined the role of discourse in supporting mathematics learning in fifth grade classrooms (Hickey, Lewison, & Mewborn, 2005). The overarching project aimed to examine the impact of regularly occurring, collaborative math activities aligned to specific state and national mathematics standards on the development of student learning. The project's goals were to support teachers' facilitation of mathematics activities and to refine activities and related assessment materials to promote conversation and collaboration, which was not normative for the schools with whom we partnered. In this study, we conduct a fine-grained interactional analysis of the how certain aspects of the designed activity structure afforded participation in math discussions in ways that differ from traditional math classrooms, which are often modeled after learning-as-acquisition models (Sfard, 2001).

We partnered with four classrooms in two midwestern primary schools for the second year of this project. Teachers of these classrooms taught all subjects, including mathematics. We worked with these four classroom teachers for the duration of the school year, throughout which we participated in monthly interventions during math class time as well as monthly meetings with all four teachers in which we reviewed data, activities, and discussed and planned upcoming activities.

## Activities

The interactions that we had with the four classrooms entailed: (1) monthly, small-group math activities in which small groups of students completed and discussed pencil-and-paper problem sets that the research team designed (three to four questions each); (2) whole class discussions led by the teacher in which representatives of each group presented problem solutions on transparencies projected overhead: (3) *conversation rubric discussions* in which groups reflected on the quality of their group discussions (the focus of our analysis, discussed below); and (4) bi-yearly pre-post formative assessments that covered similar topics to the monthly problem sets and were completed individually, were not graded, and were discussed in similar fashion to the activities already described. In addition, we also administered bi-yearly pre-post tests that served as a proxy for state and nationally administered achievement tests. The design of both the monthly activities and formative assessments included small-group feedback conversations that were meant to foster discussion of math reasoning and reflection on math discourse. We focus here on an interactional analysis of one group's participation in three of the ten monthly, small-group activities.

The problem sets in the monthly activities dealt with recently covered concepts from students' regular math curriculum, such as determining the area of polygons, prime factorization, interpreting graphic representations of data, multiplication and division of fractions, and order of operations. The sets often prioritized open-ended problems with multiple solution paths and were intended to foster extended discourse within groups. These activities were not graded, but students did receive "hints" sheets (named "Dori's Hints" after a fictional cartoon sixth grader) partway through their discussions. These "hints" were intended to scaffold possible procedures to solve the problems and to move group conversation forward.

### Conversation Rubric

In addition to the problem sets themselves and "Dori's Hints," the main focus of our analysis is the role of the *Conversation Rubric* (CR) as a mediating tool that shaped students' participation in math discussions. The CR was designed to foster groups' reflection on their math discussions about the problem sets and their positions as engaged learners within those activities (see Figure 1 below for the CR). The CR (and teacher and research team's scaffolding of groups' use of it) focused students' attention on what are considered important elements of "productive" math discourse—listening, explaining, challenging, and reflecting on understanding.

Groups worked on the CR collaboratively after discussing their reasoning and completing whole-class discussion of the problem sets. Additionally, prior to completing the problem sets on each activity day, teachers organized a whole-class discussion around the prior month's CR to encourage groups' reflection on points for improvement from last time, which usually took five to ten minutes. Our interest in highlighting the CR for examining small group discourse lies in its potential for provoking reflection on students' participation in groups. In these reflections, students were no longer talking just about math, but also talking *about* their talk about math (and sometimes taking a personal stake therein). This talk effectively created a context for students to contest the accuracy and quality of their groups' math discussions, when students actually engaged the CR, which was not always the case. We argue that this affordance of the CR evoked different ways that students could orient to their own and each other's positions with regard to the math activity, their participation in it, and what social elements were relevant to distinguish "productive" and "unproductive" engagement in math talk (Gresalfi, 2009). We argue that the traditional classroom space of received mathematical learning (Boaler & Greeno, 2000) and classroom mathematical practice was mediated by this affordance of the CR in that it prompted students to think about the ways they talk about math together—not only the mathematical procedures used, reasoning employed, or their own individual understanding, but also how productively they engaged discourse and understanding as a group.

The CR was thus a pivotal artifact (Jurow, 2005) in that it provided a shifting frame that was outside that of received mathematical knowledge, which characterizes the traditional, individualistic math instruction these students were accustomed to engaging in, because it focuses on group work and discussion. The CR was also outside the frame of the group math activities themselves, because it fostered reflection on those activities. Despite both the activity's and CR's distance from "traditional" math instruction, it is not entirely separate, as it shares some features such as pencil-and-paper format and is easily adoptable by teachers without an actual change in classroom practice. The ways that interactions afford meaning making in classrooms vary widely but are never neutral. Theoretical views on classroom talk range from characterizing it as dissemination of information to a tool for problematizing understanding, creating meaning, and positioning identity. We view interaction around the CR as a reflective meta-tool for learning on two levels—understanding how to position one's self as a type of person, and understanding ways to reason about math with others.

**Conversation Rubric**

<table>
<tr><td>Active Listening<br>Definition:<br>• Hear what everyone has to say<br>• Try to understand everyone's explanation<br>• Ask questions when you don't understand<br>• Make sure everyone has a chance to express his or her position (not necessarily in order)</td><td>How did your group do on listening?<br><br><br>What will your group work on to improve this?</td></tr>
<tr><td>Explaining<br>Definition:<br>• Describe your way of thinking about the problems<br>• Take a stance<br>• Try out new ideas by explaining to others<br>• Use math reasoning to support your answer</td><td>How did your group do on explaining?<br><br>What will your group work on to improve this?</td></tr>
<tr><td>Challenging each other<br>Definition:<br>• Be sure to voice different reasoning and solutions<br>• Draw out others' explanations with questions<br>• Agree to disagree, if necessary<br>• Everyone has something important to add<br>• Different solutions are a good thing</td><td>How did your group do on challenging?<br><br><br>What will your group work on to improve this?</td></tr>
<tr><td colspan="2">Group Reflection<br>How did your group's understanding evolve?<br>How did your group use Dori?<br>In what ways did she help?</td></tr>
</table>

Figure 1: Conversation Rubric

## Methods

The small groups in which students completed and discussed the problem sets were comprised of three to five students each, which we organized with the teacher at the beginning of the school year based on mixed student math ability and parity across male and female students. In each of the four classes with whom we partnered we video-recorded two focal groups (limited due to equipment and personnel constraints), which we selected based on teachers' recommendations about likely talkativeness (a positive thing when examining discourse), consent to be video-recorded, and diversity of socioeconomic status, race, and math ability across the eight groups.

The data for this study include three, hour-long video-recorded activities of a four-student focal group comprised of two boys (Trey and Adam) and two girls (Christy and Mandy; all names are pseudonyms). We chose this group because they remained intact for all but the first activity day of the year, and they often talked at length about the CR. We chose these three activity days—day 2, day 4, and day 10—out of the total ten because they included complete video records of the CR discussion that also represents a range across the entire group trajectory throughout the year. Our analysis examines the types of positioning and learning that the CR fostered, both in the discussions around the problem sets and in the CR discussions themselves. Since the CR was discussed at the start of each activity in a whole-class discussion led by the teacher, we found that it shaped elements of discussion throughout each activity day. We use discourse analysis informed by sociocultural and sociolinguistic commitments (e.g., Anderson, 2009; Bucholtz & Hall, 2005; Mercer, 2004), including the immediate details of talk (pauses, pronoun usage, indexicals, content) as they relate to social positioning as well as categorizations of the group's trajectory of participation across the ten months of the intervention.

## Findings

We present the overall findings in terms of the elements of the CR for each episode. Our analysis includes both verbatim snippets from the group's CR discussion and our commentary on how that aligned with what we saw actually happening in the video recordings of their math discussions preceding the CR discussions. We first present an account of how the students used the conversation rubric and the alignment between their reflections on their participation and our observations of their participation. We then consider the episodes in terms of positioning.

## Day 2

The group's reflections on their activity from the second day in the CR discussion was generally positive; for the most part, the group stated that they had collaborated quite productively as a group, at least with respect to the items to which they responded. Specifically, they claimed that: "we tried to listen to [the explanations]"; "[we] heard what everyone had to say," and "'sort of'" asked questions when we didn't understand each other." Likewise, the group claimed that they had offered productive explanations, stating that they had "half and half tried to describe our way of thinking about the problems." "We said 'I don't think that's right' if we didn't understand each other," and "did not just sit there and nod our head." "We used math reasoning." In contrast however, they noted that they did not robustly challenge one another when working on the problems, stating: "We didn't really voice different reasoning and solutions except at the very end (after Dori)." Although one student claimed that their performance at challenging one another was only mediocre, others argued they did well because "we actually tried and did our best."

These reflections were generally not aligned with our observations of students' actual math discussions on Day 2: although students did at times listen to each other to the extent that they shared answers or procedures, they did not work collaboratively beyond "reporting out." As an example, after the students had finished the first problem on their worksheet, they were asked to share their answers with their group. The following exchange ensued:[1]

**Christy:** Um, I did drawings like 12.5 plus 5, you just add. You just, like, add. So I just added 5 to that one. And then if you do, um, 12.5 times 5, you'd be doubling, the number over and over again. So, like that's why I thought it would be greater. Because with this you only add 5 and you'd be doubled.
**Trey:** I didn't get anything.
**Christy:** Well, how far did you get?
**Trey:** I didn't get any [inaudible] [has covered his face with his paper]
**Christy**: I thought you had something wrote down.
**Trey:** But then I erased it.
**Christy**: [looks at Adam who is concentrating on his paper] Well, what did you think? [looks back at Trey]
**Trey:** [closes the paper closer over his mouth] My mind is blank.
**Mandy:** Wait, what?
**Trey:** [turns head to Mandy] My mind's blank.
**Christy:** OK. Adam, what did you get?

This exchange represents an "explanation relevant" moment; one of the members of the group expressed confusion about the problem but did not receive an explanation or assistance from the rest of the group. In this way, Trey was positioned as both off task and unworthy of assistance, as his confusion did not merit discussion or explanation from any of the other group members.

## Day 4

The group began this day prepared to listen actively to each other, because they decided that "we weren't listening to each other's problems last time; We're gonna try to listen to each other more," in their discussion about what to improve upon this day (based on last months' CR). Importantly, the group left their goal at this relatively unspecified level; they did not detail what listening actively looked like, or what specific steps they would take to ensure that they were realizing their goal. In reflecting on their discussion of the problem set (completed in two pairs this time) using the conversation rubric, the students did not evaluate their listening very positively; Christy said that she and Trey "did really good listening to each other and asked questions and communicating," but Mandy reported that she and Adam were only "sort of" listening to each other at the end a little bit. Interestingly, the students did not refer back to their stated goal for the day. With respect to explaining, Adam said (of he and Mandy), that "we did good because at least we explained. It may have been wrong but at least we challenged with respect and stuff." However, as a whole the group concluded that they could have done more to challenge each other because they didn't talk or listen all the way through.

Again, our observations of the group's work did not conform closely with the students' own characterization of their work. The group broke into pairs (as instructed by the teacher). Christy and Trey worked mostly independently next to each other; they listened to each other's answers and reviewed their procedures, but mostly they did not engage each other beyond stating their answers and reasoning. They did not actually work on the problem together. Mandy and Adam, however, did not even discuss what they

[1] Transcription conventions: [text] = description of non-verbal action; TEXT = stressed speech.

had written with each other and required extensive intervention from one researcher to even comparing their answers.

### Day 10

Adam was absent this day, and although the teacher and research team encouraged the rest to work as a threesome on the problems, Trey resisted (and Mandy immediately said that was fine with her and Christy). This final day of our analysis illustrates the extent to which the group's interaction came to a head; whereas in the previous two episodes, reflection about their work was generally not contentious and only tangentially related to the group's work, in this final episode the CR became an explicit tool for positioning each other in relation to the group's work. Specifically, although Christy and Mandy claimed that they had worked well together, Trey argued vociferously that they had neither listened nor explained *to him*. In our observations of the video data, although Christy and Mandy worked together without conflict, their collaboration was rather one-sided with Christy leading the work and Mandy writing answers and neither asking for, nor receiving, explanations. Trey worked independently, initially by his own desire, and later when the three were meant to discuss what they had written because he could not get the attention of the rest of the group.

In their discussion of their collaborative work, the CR became a mediating tool that Trey leveraged to successfully note his dissatisfaction with the group's discussion (or lack thereof). Although he had previously attempted to get the attention of the two girls, it wasn't until the CR became a topic of discussion that he was able to engage his group mates productively.

**Christy**: Listening? We listened to each other.
**Mandy**: What, where?
**Christy**: We listened to each other [points to CR].
**Trey**: No we're not done yet. We're not done yet [meaning not ready to discuss the CR].
**Christy**: Too bad.
**Trey**: You can't do it [CR], we all have to do this together [discuss their work].
**Christy**: [to Trey] Do you think we listened good?
**Trey**: [pauses] No.
**Mandy**: Yes.
**Trey**: NO, we haven't listened at all to each other.
**Christy**: That's because you didn't want to work with us, so too bad. We listened to each other!
**Trey**: But you cut- we all have to agree!
**Mandy**: But you said you didn't want to work with us.
**Christy**: No, you just don't want to work with us, so it's [inaudible]
**Trey**: No but we all have to agree. She said when you're done after we discuss this [lifts up activity sheet. Christy keeps writing on CR with Mandy looking on. 16 sec pass. Trey picks up their written work and examines it].

As the excerpts and descriptions above illustrate, the group did not really engage the CR in deep and meaningful ways, which was the case across the 10 activity days throughout the school year. For this reason, we argue that although they clearly used the CR for interpersonal positioning, especially on day ten, they did not engage in disciplinary positioning around the CR discussion (one of our hopes in designing it).

## Conclusions and Implications

To revisit our earlier stated aims, in this paper we considered one small student group's discourse across three monthly, collaborative math activities as evidence of the extent to which the *Conversation Rubric* was a mediating tool that supported collaborative engagement in math activities and shaped students' reflection about their understanding and their positions in discussions in ways that might not have occurred without the CR. With respect to these two points, we conclude from our interactional analysis of the three activity days that the CR supported competitive rather than collaborative engagement for this group, and that it failed to shape actual reflection on math understandings. We base this claim on observance that the group tended to focus on procedural display, task completion, and compliance with teacher formulations of the CR discussion as well as arbitrary group-dictated rules rather than considering how they might learn from each other or reflect on changes in understanding about the math from beginning to end of each activity day (or between activity days—another goal of the CR that was not realized for this group.

Clearly the CR had affordances for shaping group discussion, but how these affordances were realized as compared to how they were envisioned by the research team, at least with respect to the evidence supported in this case study, begs the question of whether the CR in fact does not afford reflexivity, or whether its use in this classroom with this particular group simply thwarted these affordances. In terms of design, we feel that if the group had taken up the CR as an actual platform for considering their math discussion, it could have worked. However, this group became fixated on either getting it done quickly and complying as if it were a set of boxes to tick, or using it to enforce compliance with dominant group members' whims (as seemed to be the case on day 10 with Christy and Mandy). As we stated above, activity designs can afford productive tensions between students that lead to positioning of each other. In this case, we observed tensions, but they unfortunately were not productive in terms of pushing understanding and learning from difference. Instances of different mathematical representations for this group were either glossed over as the same representations (so they could move more quickly through the activity) or became arguments about who was "right." As a result, the group never moved beyond initial understandings of the math activities, and they never reflected on these differences in the CR discussions. To further explore the design implications of the CR, we will be analyzing all eight groups' discourse and positioning in a future paper, of which this analysis will serve as a model for the analytic methods. In addition, we will also compare the second year iteration with that of the first year's groups' work on the CR, which was slightly different in form and enactment (see Anderson, 2009 for an in-depth analysis of one group's positioning relative to the CR in the first year).

Our methodological contribution from this analysis includes considering how a mediating tool shaped the texture of learning across small-group discussions, which we examined across time. In this student group's case, the CR-as-artifact shaped math discussions but not in ways we had intended or hoped. The CR-as-methodological lens, however, provides us with useful implications for considering the role of math discourse in supporting opportunities to learn, as we can learn considerably from negative cases. We examined the CR discussions as discrete points along the group's trajectory where values became explicitly associated with participation explicitly in light of intervening math discussions. In this way, the CR foregrounds how participation in small group discussion highlights tensions surrounding accountability—*especially* when such practice is not the norm for a classroom.

This analytic lens on the surfacing and navigation of tensions surrounding group participation is useful for understanding how interpersonal positioning can afford or thwart deep math discussions. For this group, their trajectory across the ten activity days spanning ten months illuminated a practice that did not change—their explicit attention was always focused on listening, but they never really changed how they listened to each other much beyond the second day. In terms of learning, it is unclear whether the math discussions or the CR aided this group's collective or individual understanding. We argue that these lacks stem from a failure on the group's part to get beyond the notion of their math work as their own, in other words something to which they alone were individually accountable. The group treated the math activities as their own work for which they would eventually be evaluated and the CR as a rote activity in which to give the appearance that they were doing things as a group. However, because the CR was not pushed or taken up in this classroom in a meaningful way, it thus failed to be meaningful. From these conclusions we summarize that interpersonal positioning must be navigated so that it does not get in the way of engaging the math, as it did for this group because vulnerability and righteousness clouded their ability to see the math for the spats. Instead, interpersonal positioning must lead into an appreciation for the variety of different ways there are to approach the representation and solving of math problems. In this way, the group we considered here "talked with their mouths full" of the CR, in that they interpreted their participation through its lens. However, that mediating tool only shaped participation after mathematical activity was completed, and thus the CR served more as a snack than a meal.

## References


Anderson, K.T. (2009). Applying positioning theory to analysis of classroom interactions: Mediating micro-identities, macro-kinds, and ideologies of knowing. L*inguistics and Education, 20(4), 291-310.*

Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education, 19,* 118-126.

Ben-Ari, R. (1997). Complex instruction and cognitive development. In E. G. Cohen & R. A. Lotan (Eds.), *Working for equity in classrooms: Sociological theory in practice*. New York: Teachers College Press.

Boaler, J. & Greeno, J. (2000). Identity, agency, and knowing in mathematics worlds. In Boaler, J. (Ed.). *Multiple perspectives on mathematics teaching and learning* (pp. 171-200). Westport, CT: Ablex.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside school. *Teachers College Record, 110*, 608-645.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A Sociocultural linguistic approach. *Discourse Studies, 7,* 585-614.

Cohen, E. G., & Lotan, R. A. (1995). Producing equal status interactions in the heterogeneous classroom. *American Educational Research Journal, 32(1)*, 99-120.

Davies, B., & Harré, R. (1999). Positioning and personhood. In R. Harre & L. V. Langenhove (Eds.), *Positioning theory: Moral contexts of intentional action* (pp. 32-52). Oxford: Blackwell.
Gresalfi, M. S. (2009). Taking up opportunities to learn: Constructing dispositions in mathematics classrooms. *Journal of the Learning Sciences, 18*, 327-369.
Gresalfi, M. S., Martin, T., Hand, V., & Greeno, J. G. (2009). Constructing competence: An analysis of students' participation in the activity system of mathematics classrooms. *Educational Studies in Mathematics, 70,* 49-70.
Hickey, D. T., Lewison, M., & Mewborn, D. (2005). Multi-level assessment for enhancing mathematical discourse, curriculum, and achievement in diverse elementary school classrooms. National Science Foundation Grant (REC 0440261).
Holland, D., Skinner, D., Lachicotte, W., Jr., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.
Horn, I. S. (2005). Learning on the job: A situated account of teacher learning in high school mathematics departments. *Cognition & Instruction, 23*, 207-236.
Jurow, A. S. (2005). Shifting engagements in figured worlds: Middle school mathematics' students' participation in an architectural design project. *The Journal of the Learning Sciences, 14,* 35-67.
Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., et al. (2007). Specifying computer-supported collaboration scrip*ts. Computer-Supported Collaborative Learnin*g*, 2,* 211-224.
Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*, 29-63.
Mercer, N. (2004). Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics, 1,* 137-168.
O'Connor, M. C., & Michaels, S. (1996). Shifting participant frameworks: Orchestrating thinking practices in group discussions. In D. Hicks (Ed.), *Discourse, learning, and schooling* (pp. 63-103). New York: Cambridge University Press.
Sfard, A. (2001). There is more to discourse than meets the ears: Looking at thinking as communicating to learn more about mathematical learning. *Educational Studies in Mathematics, 46,* 13-57.
Staples, M. (2008). Promoting student collaboration in a detracked, heterogeneous secondary mathematics classroom. *Journal of Mathematics Teacher Education, 11*, 349-371.
van Langenhove, L., & Harré, R. (1999). Introducing positioning theory. In R. Harre & L. van Langenhove (Eds.), *Positioning theory: Moral contexts of intentional action* (pp. 14-31). Oxford: Blackwell.
Wortham, S. (2004). The interdependence of social identification and learning. *American Educational Research Journal, 41*, 715-750.

## Acknowledgments

We would like to thank the PIs on the grant that supported this work, Dan Hickey, Mitzi Lewison, and Denise Mewborn as well as the National Science Foundation without which this data could not have been collected. We are also grateful to other members of the research team—Dionne Cross, Asli Ersoz, Peiwei Li, Ingrid Graves, Lenny Sanchez, and Erik Winarski.

# Fostering meaningful scientific argumentation practices through ongoing classroom interactions

Xiaowei Tang, Janet Coffey, University of Maryland, College Park, 2226 Benjamin Bldg, UMD, MD20742
Email: kitty@umd.edu, jecoffey@umd.edu

**Abstract:** Specific sets of structural elements are often taken as the criteria for assessing students' argumentative abilities and designing curriculum or teaching strategies that promote classroom argumentations. In this paper, we take issue with such oversimplified, research paradigm, arguing for the necessity of examining students' argumentative practices as rooted in the complex, evolving system of classroom. Employing a sociocultural-historical lens, we illustrated through close discourse analysis how a high school biology class continuously builds up affordances and constraints for their argumentation practices through interactions. Whether argumentative conversations can take place in certain situations and sustain, and how the teacher and students participate in it have much to do with the learning goals, norms, teacher-student relationships and epistemic stances constructed overtime. Based on such findings, we suggested that the field should consider promoting classroom scientific argumentation as a long-term process, requiring supportive resources to be developed through continuous interactions.

## Argumentation: Structure, Skills and Classroom Norm

Over the last half century, scholarly views of science have shifted from a dominant focus on experimental processes to increasing concerns about its socially constructed nature (Driver, Leach, Millar & Scott, 1996). Conceiving science as ongoing human activities, philosophers of science have suggested scientific argumentation as playing a central role in the production of scientific knowledge, constituting a core element of scientists' practices (Latour & Woolgar, 1986; Fuller, 1997). Following in suit, reforms in science education has made efforts to promote scientific argumentation practices (NRC, 1996, 2000; Kelly, Druker, & Chen, 1998; Driver, Newton, & Osborn, 2000).

A framework commonly used by educational researchers to conceptualize argumentation derives from Toulmin's (1958) *Use of Arguments* (Bell & Linn, 2000; Driver, et al, 2000). While the original purpose of Toulmin's work was to challenge the traditional, inference-centered view of argument with practical emphasis on justification, it was reduced to a model consisting of the typical structural elements of argument and their functional relationships. In short, *data* refered to the evidence used to support a *claim*—the point-making statement, and warrants were the logical statements bridging *data* and *claim*. (Other elements of his framework -*qualifier*, *rebuttal* and *backing* –are less frequently employed in educational research).

This general account of formal argument has been modified into important tools for examining and evaluating argumentation practices in classroom (Kelly, et al, 1998; Jime´nez-Aleixandre, Bugallo-Rodrı´guez, & Duschl, 2000; Erduran, Simon, & Osborne, 2004; Dawson & Venville, 2009). Some directly applied a Toulmian model as the coding scheme, analyzing argumentation discourse for how frequently specific structural components got involved (Kelly, et al, 1998; Jime´nez-Aleixandre, et al, 2000); others developed scaling measurements, assessing student arguments based on structural completeness (Erduran, Simon, & Osborne, 2004; Dawson & Venville, 2009). With similar emphasis, early work from Kuhn (1991, 1993), though not employing a Toulmian framework, also linked argument structure with cognitive skills in coordinating theoretical claims and evidence.

Findings from research focusing on argumentation structure point to both the incomplete structure of students' arguments and the weaknesses in students' argumentation skills. For example, many found that students often argued by stating their claims without justification through warrants and data (Jime´nez-Aleixandre, et al, 2000; Dawson & Venville, 2009). Through interview studies, Kuhn (1991, 1993) suggested that adolescents and lay adults lack the skills to distinguish data and theoretical explanation of data. These deficiency explanations motivated the development of specific curriculum and instructional strategies for initiating, supporting and scaffolding students' argumentation practices, such as explicit teaching of general reasoning patterns (Zohar and Nemet, 2002), using open-ended argumentation prompts (Erduran, Simon, & Osborn, 2004), creating opportunities for peer or small group discussions (Erduran, Simon, & Osborn, 2004), and replacing written scaffolds with teacher introduction of an explanation framework (McNeill, Lizotte, Krajcik, & Marx, 2006). Through pre and post assessments designed to measure the structural completeness or the usage frequency of certain components, these interventions showed satisfactory changes in students' argumentation discourses.

While this research proves prevalent, its guiding framework and overarching goal greatly limits the scope of understanding on classroom scientific argumentations. As a structural account of formal arguments, Toulminian scheme provides neither tools nor language for analyzing argumentation as a goal-driven, socially constructed discourse phenomenon, embedded in either classroom discourse or the broader discourse of science (Kelly, et al 1998; Driver, et al, 2000). Judging the quality of a scientific argument according to structural features, in separation from its functional role and the cultural-historical contexts in which it is rooted, can be misleading. First, as Kelly et al (1998) suggested, language can be flexibly used in conversation, and the need for justification is often shaped by the interactive history and shared knowledge. He noted it was difficult to identify components of an argument without considering the related contexts and conversational dynamics, and even more difficult to make fair inference on individuals' abilities or skills in argumentation based on the structure of their arguments in an interactive discourse.

Second, the value of argumentation in science rests in its power to resolve scientific discrepancy (Driver, et al, 2000). If argumentation skills get foregrounded and established as the learning goal, but learners are not pursuing something that would raise the need to argue, we have little to ensure that argumentation gets developed as a useful discourse tool that students can draw on, and risk making parts of an argument another school-defined convention.

With such concerns, some scholars suggested that scientific argumentation should be established as a classroom norm (Driver et al, 2000; Engle & Conant, 2002). They called for argumentation to become evident in discourse patterns that permeate a variety of classroom activities, emerging naturally as part of students' regular sense-making interactions. In alignment, recent research has begun to investigate the complexity of classroom argumentation phenomena, taking into account multiple, interdependent factors of the learning environments. Studies on how learning contexts motivate engagement in argumentation suggested the important roles of factors such as social norms of classroom interactions (Berland & Reiser, 2009) and epistemological resources (Louca, Elby, & Hammer, 2004). In a case study of an emergent and lasting classroom argumentation, Engle and Conant (2002) demonstrated how a class' disciplinary engagement was constructed through continuous effort in fostering a learning community that encompassed such features as attitude and values towards controversial ideas, positions and roles of students, accountability for disciplinary norms, and supportive classroom resources. This line of research conveys a resounding message: to understand classroom affordance of meaningful argumentation practices, we need to look into the non-linear, multifaceted development of classroom learning environments.

Following this train of thought, our study explores how classroom norms conducive to scientific argumentation take shape and give rise to students' sense making interactions in a high school science classroom. From a sociocultural-historical perspective (Vygotsky, 1978; Engestrom, 1987), we view learning as continuously occurring in interactions between the social and individual planes, mediated by artifacts and sociocultural contexts and developed through the interactive history of the specific group and broader communities. From there we ask these questions:

*What affords or constrains the use of scientific argumentation in science classrooms?*
*How do classroom affordances and constraints of argumentation get constructed over time through classroom interactions?*

## The lens of localized activity theory

The dynamic and evolving view of classroom activities came from the theoretical lens of Activity Theory (Engestrom, 1987, 1999). According to this model, in a continuously changing activity system, collected subjects conduct object-driven activities that provide the momentum for its changes. Their actions and interactions towards the object are mediated by material and symbolic artifacts as well as factors from the broader contexts such as rules, community and labor division. Localizing this framework to fit the specific features of my subject, we consider a classroom as an activity system in which teacher and students carry out activities driven by overarching and specific learning objects, or other classroom/school related objects (such getting familiar with each other, setting dates for exam). These activities are conducted employing material tools (such as textbooks) and communicative tools (such as language). Contextually, activities are also shaped by norms and routines of a classroom (rules), values and norms held by the educational system (community), as well as the roles of and the relationship between teacher and students (labor division). I referred to all these factors as *classroom mediational resources*.

The term "resource" has already gained various meanings within the field: Hammer (2000) employed it to describe the conceptual and epistemological "knowledge in pieces" (diSessa, 1988) brought to learning by individual students. Crawford, Kelly, and Brown(2000) used it to refer both to individual knowledge of science practice and the artifacts affording science practice; Engle and Conant (2002) broadly defined "resources" as supporting settings ranging from available space of class time to classroom constructed artifacts and norms to modes of discussion. In this study, the meaning of *resource* is closer to Engle and Conant's (2002) use, but broader (e.g.: to include epistemological resources), and focuses on a functional common ground, that is, when functioning together, they afford certain interactions and constrain others.

Mediational resources get constructed in interaction, both explicitly and implicitly. Explicit construction occurs through actions/interaction with clear goals of setting up resources for current and/or future learning activities. Such situations, do not appear as common as implicit constructions, in which the interactions are not focusing on setting up resources, but as certain discourse and participation patterns, compatible values, norms and relationship that emerge and hold potential to mediate current and future interactions.

Within this framework, scientific argumentation is a general form of interaction serving the objects of different classroom activities. For such interactions to emerge, the class needs to develop shared understandings of when argumentation is allowed, what type of language to use and how to participate. The norm of argumentation also needs to be compatible with other classroom norms, both content wise and context wise. To illustrate the establishment of such argumentation "norm," we have to explore the complexity of an evolving classroom system. The influence of mediational resources for argumentation extends beyond structural concerns to content, purpose, values and participation patterns of specific argumentation interactions.

## Data Source and Methodology

To investigate classroom affordances and constraints on scientific argumentation, we videotaped a teacher's (Sarah's) class for a full semester (September 2008- January 2009). The class was an ethnically diverse, honor's level, biology class, comprised of 24 students, with a roughly balanced gender ratio and a mix of 9$^{th}$, 10$^{th}$ and 11$^{th}$ grade students. At the time of this study, Sarah was in her 3rd year teaching biology in a large, suburban-urban fringe public high school. She had participated in a university-based collaborative research project that sought to better understand teachers' curricular modifications. During that project, she was identified by researchers as relatively good at scaffolding student argumentation and as a teacher who strove to facilitate students' participation in other scientific inquiry practices, despite institutional pressures associated with standardized curriculum guides and high-stakes testing that could focus attention elsewhere.

Our analysis of classroom data focused on interactive learning discourses, mainly speech, but including contextualization cues (Gumperz, 1982; Duranti & Goodwin, 1992) such as eye gaze, gesture, writing, and visual representation on the board. Initial analysis entailed cataloging summaries of interaction content of all classroom video, in one to five minute intervals. The thickness of descriptions (Geertz, 1973) was determined by the richness of the discourse data and how closely it related to the interest of this research—thicker where students shared ideas and built arguments.

We began to identify discourse patterns and shifts on different levels: 1) General discourse structure, such as triadic dialogue (Lemke, 1990), reflective toss (van Zee & Minstrell, 1997), and back-and-forth argumentation (Cazden, 1988); 2) Ways in which productive learning interactions were initiated and maintained, which included type and amount of teacher scaffolding; 3) (i.e.: scientific language and everyday language (Lemke, 1990)); and 4) Patterns of expressions (common words or phrases).

Episodes bearing patterns or demonstrating shifts in patterns were transcribed and subjected to closer analysis in order to draw out meanings of discourses and gain insight into the social processes underlying the learning interactions. (Gee, 2000; Erickson, 2004; Kelly, 2005). To do this, we summarized participation and discourse features of each episode. We then worked backwards (and forwards) to try to identify an interactive history through which parallel features emerged. Analyses across episodes helped bring out possible meditational resources, how they were constructed and what work they accomplished for classroom learning.

## Data Analysis

To illustrate how such theoretical and methodological frameworks guide our analysis, we draw on several representative argumentation episodes from the class we studied.

### An Illustration of the Classroom Argumentation "Norm"

The first one took place during a class review of biotic relationships (predation; commensalism; mutualism; parasitism), which is a month into the fall semester and a week after they took notes on that topic. The teacher Sarah asked students to name and explain the meanings of different biotic relationships. After she confirmed that parasites usually do not kill the host, one student threw in a challenging question and a discussion ensued:

Acer: if there is like a parasite or something in the prey or whatever dies, like, because of it, would you call it predator or what or still parasite?"
Charles: like you could die from mosquitoes… like the disease they carried...
Sarah: so the disease they carried like kills them, not the mosquito.
Charles: well, the mosquito bit the person.
Students: Yeah…
Christine: Yeah, but it wasn't/
Sarah: but the bite didn't kill the person.
Christine: It is the disease…
Acer: Um, What, wha…isn't a virus like a parasite?
Charles: but they inject the virus that caused the disease.
Sarah: So, here is, if I were gonna argue back with Acer. He just said isn't a virus like a parasite. If I'm gonna argue back…Does anybody else wanna argue back?
Students: No.
Sarah: I'm gonna argue that/
Charles: Because virus isn't an organism.
Sarah: some people would not consi…virus is kind of falling in this iffy place where we are not sure we should call them alive or not alive. So I would argue that is not a relationship between two living organisms. I would say that there are always exceptions to the rule, like, Tim said, "That's rare." There are exceptions though. So usually parasites don't kill their host. If they did kill them, I don't know, maybe you would want to call it a predator 'cause it hunts it and kills it.
Tim: but usually it just doesn't happen
Tim: like it would be predator and prey if that in that form. [inaudible]
Sarah: what did you say?
happens a lot, but it can happen sometimes but usually the other animal doesn't die, so it is just parasitism.
Charles: Oh, my gosh, he sounds so smart. [class laughter]

This collaboratively constructed scientific argumentation flew through shorthanded and co-constructed arguments, demonstrating the students' ability to engage in scientific argumentation and also reflecting their conceptual understanding of the subject. As Acer problematized the demarcation between parasitism and predation with a particular situation—a parasite in a prey that dies from it, he made an argument based on a hypothetical overlap between the two categories. This required certain conceptual understanding of both types of relationships. . The exchanges between the teacher (and Christine) and Charles showed that they both have certain understanding of the pathological mechanism behind mosquito-borne, transmissible disease, but have different interpretation of the case in terms of what is the killer: while the teacher identified the disease rather than the mosquito as the direct cause of death, Charles defended his argument by emphasizing the mosquito's initiator position in the causal chain—it "bit the person" and "injects the virus." The argumentation shifted when Acer brought up virus as an example for the situation he described. Charles and the teacher both argued through denying virus' status of living organism and excluding it from the category of parasite, which is the precondition for it to be considering as a candidate for the situation they talked about. By introducing the "iffy place" about virus' being alive or not, Sarah indicated the debatable nature of her rebuttal, and also took a step back, acknowledging the possibility of a loose boundary—"If they did kill them," "maybe you would want to call it a predator 'cause it hunts it and kills it." Tim then jump in and showed an understanding of the probabilistic meanings behind accident and normality

While this episode shows merits if examined in terms of argument structure or content, its significance extends far beyond that. First, the argumentation is initiated by a student's challenge of the canonical boundary between parasitism and predation, which had been established by the teacher in previous class sessions and confirmed just prior to this episode in a teacher-student interaction. Such moves are rare, in our observations and in the common patterns of classroom argumentation interactions recorded by the literature, as usually argumentations get start with open question and without "correct answer" being given (Engle, 2002). Second, this conversation occurred during a review activity, where checking conceptual understandings and knowledge of terminology is far more typical than engaging in scientific argumentation. Third, the students brought up arguments on their own rather than following specific probes from the teacher. While the teacher's scaffolding moves were very limited (The main move she made is bringing Acer's argument to the class' attention and ask "Does anybody else wanna argue back?" which is followed by the "virus is not alive" argument co-made by Charles and herself.), the students attended to the ideas from each other and from the teacher. Lastly, the nature of the teacher's participation is distinct from more commonly adopted roles: she was an arguer in the field, actively constructing her own argument. Though some might find some authoritative flavor in her arguments, we noticed that the students treated her ideas on par with their peers' arguments rather than as ones to accept. They argued with her without hesitate, clearly demonstrating their understandings and critical thinking.

We argue that understanding the features listed above are important for explaining how argumentation took place at the moment. We would also argue that these are the indicators that certain argumentation "norm" has established in this class, so that argumentation can naturally emerge and contribute to the class' sense-making process. The immediate contexts of this episode, however, does not explain what gave rise to these features. When examining this episode as embedded in the interactive history of classroom, we have identified previous episodes that share or contrast this episode in certain discourse features. Through comparing multiple episodes, we analyze how mediational resources conducive to scientific argumentation get constructed, and contribute to the features and participation patterns seen here. The following section demonstrated the construction of one piece of such resources.

## "Because You Keep on Asking"

While the students attend to the substance of each other's and the teacher's ideas in the parasitism-predation argumentation, this was not always the way they interacted in this class. On the third day of class, in a discussion on "why do we brew tea in hot water?" after the students came to agreement that hot water works faster than cold water since molecules move faster in it, Sarah posed the question: "At the end of all, are we going to get the same product using hot water and cold water?" At first, the class together made the "yeah" sound in a low-pitch, elongated tone, as if that answer was so self-evident. After this question being pushed back twice, only several students still remained in responding, and their answers provided a simple causal explanation for their previous judgment —"it (the tea) would eventually go into water." As Sarah, again, pushed it back to students by explicitly asking for "anybody thinks no?" the following exchanges took place:

Sarah: Does anybody think no, we are not gonna get the same thing?
Dani: (raises up hand)
Sarah: Yeah? How come?
Dani: Because I think that's the right answer.
Sarah: And … why?
Dani: Because you keep on like, asking that.
Sarah: I'm just trying to get you guys disagree about stuff! One person said something and you were all like, yeah— (she mimicked their lazy tone. The whole class laughed and Tristan put his hand up).

In response to Sarah's push, Dani raised his hand and voted for the "not gonna get the same thing" claim. When Sarah probed for his reason by asking "how come?" he gave no scientific explanation but stated that he thought "that's the right answer." It was after another push from Sarah he shared his interpretation that the teacher's keeping asking when everybody said "yes" must suggest the opposite to be correct. Danny did not receive any counterargument, laughter, or other looks from his classmates; instead, the class remained silent with their eyes fixing on Sarah, implying that they were also waiting for an answer. Sarah did not show much surprise either. Claiming the purpose of "get you guys disagree about stuff," she then imitated and made fun of the way students casually threw away their agreements, indicating that she would interpret such answers as not resulting from critical thinking.

At this moment of the discussion, students did not generate scientific argument. Even the teacher's move to scaffold was misinterpreted as an indicator of correctness. Sarah did not play the arguer role here. Since the students would accept what they thought she indicated, it is reasonable to guess that they were concentrating on getting the correct answer and probably would not stand on equal foot and argue with her.

This conversation did not end there. Students offered an alternative answer, and one student provided a piece of reasoning that would support the claim: there would be less water in tea made with hot water due to evaporation, so there would be a higher concentration of tea. This argument was not taken seriously and respectfully, as the class immediately exploded in laughter and made funny "ohhh" soundand one student even sarcastically stated, "what an awesome answer" with rolling eyes. Sarah, however, ignored the sarcasm and grabbed the idea. She offered the student a chance to repeat his idea to the class and then encouraged him to further his chain of reasoning.

This discussion ended up with several students claiming that hot tea resulted in a higher tea concentration. Other classes of Sarah's ended up with the opposite conclusions. As Sarah told us in the interview, there was "not an exact right answer such discussion tried to reach." Reasoning-based arguments on either side would count towards fulfilling the object of the activity.

In contrast to the student-initiated predation and parasitism-predation argumentation, although this tea brewery discussion was planned to be an open-ended one, the way the students participated indicated their expectation of a correct answer from the teacher. Instead of thinking critically about the phenomenon or reasoning through ideas, the students pursued the teacher's confirmation. Sarah made much effort scaffolding: first she encouraged argumentation by pushing back reasoning-lacking agreement; she probed students for reasoning, and communicated the purpose of her discursive move; finally, by requesting students to articulate a chain of reasoning, she claimed value for this spark of scientific thinking, and set an example for the class on how to focus on the scientific substance of ideas.

When contextualized in the larger picture of school education and the settings of this classroom, what gets communicated here is much more than what has been explicitly said. By explicitly associating her discursive move with promoting disagreement, and by attending to the scientific substance of an idea the class laughed at, Sarah also negotiated with the class a shift in the general goal and values of science learning. This shift, we would argue, is a piece of resource that could contribute to the ways in which students attended to ideas in later argumentation (such as in the parasitism-predation discussion).

We have identified other resources through comparative analysis of episodes of classroom interactions. Due to the page limit, we summarize them in the findings but would not go into details here.

### Constraints on Classroom Argumentation

While Sarah's class has constructed significant affordances for student participation in scientific argumentation, it is not the case that scientific argumentations can be developed towards any direction at any time. We saw constraints on argumentation in that class. For example, the parasitism-predation argumentation lasted for a few minutes and did not develop to its fullest. The conversation was called to an end when students started to challenge the "iffy status" of virus. Sarah's decision to "get back on topic" highlights that not all tangents are fair game, and not all argumentation are permitted. The predation-paratism exchange showed how the class was afforded to go on tangents, initiating and sustaining argumentation in an activity not particularly framed for scientific debate or open-ended scientific discussion. In that part, the arguments focused on demarcating the concepts of parasitism and predation. However, with tighter objectives of a review activity's object on understanding biotic relationships, redirecting the conversation was not an option. In previous classes, we have observed situations when similar tangential argument got accepted and followed in activity with looser and multifaceted objects (not shown here): for one thing, the activity was set up as a chance for everyone to participate, attending to, challenging and building on each other's ideas; for another, they were trying to make sense of a truly unsolved scenario, which even the teacher was not sure what was going on. This allowed (and actually required) in depth explorations into the details of ideas. In contrast, the unit review activity has a much more specific goal and therefore clearer boundary for tangent, since there was a more to-the-purpose topic to "get back on".

How, then, should we understand such constraints? What contribute to where and when they present? What determine the strength of their effects on ongoing conversations? When tracing back the class' interactive history, we identified mediational resources constructed by the class that would contribute to such constraints.

## Research Findings

The major finding of this research, as partially depicted in the analysis above, is a more contextualized account of classroom scientific argumentation. We go beyond evaluating and boosting certain argumentation structures, moving into the construction and functioning of argumentation-related classroom mediational resources. Our analysis demonstrates how classroom norms, rules, teacher-student relationship and discourse patterns formed through interactions, some of which even seem unrelated to argumentation, can later mediate students' participation in argumentation.

Through analysis, we have identified resources that could organize and mediate the class' participation in scientific argumentation in activities not framed as ones for practicing argumentation. These include: A comfortable-to-share and whole-person-respected classroom relationship; general goal and values assigned to the substance of ideas; the participating norm of "talk about shared ideas;" and repeatedly conveyed epistemic messages that biology concepts "always have exceptions."

We have also noted the tenuous nature of the classroom argumentation "norm" in the face of system-wide presses, like common curricula and high-stakes testing. We identified interactive history the discourse constructs of "fun" and "serious/boring," which set different participation norms for activities focusing on substance of scientific ideas and activities focusing on test preparation. As unit exams and the final state assessment became more impending, we saw significant changes in the teacher and students' participation patterns, including a drop in the frequency and length of student-initiated argumentations. Some other constraining resources included the different values and different nature of knowing attached to different biological subdomains and the classroom norm for resolving goal conflicts.

## Research Significance and Implications

We chose to write about this classroom, not just for the distinct argumentations it affords, but also for the weight of realness it bears. The word "realness" here has two folds of meanings. One points to the systemic pressure this classroom has to deal with—since classrooms cannot always be experimental sites in vacuums. In order for the goal of scientific argumentation to organize classroom learning in a meaningful way, it has to coordinate with many other objects that orient classroom activities. The other points to the goal of promoting argumentation in science education: to construct a useful way that affords students' sense making process, rather than as a performance to satisfy researchers' criteria, then researchers, need to better situate the practices of argumentation in the broader fabric of schooling and science.

This study aims to capture how classroom argumentation "norm" grows out of (or fades away through) sequential classroom interactions. As the story of Sarah's classroom unfolds and as our analysis goes deeper and more thorough, we identified mediational resources that contribute to the affordance and the constraints of argumentation practices in this classroom. This provides a reference of things to account for in establishing the classroom argumentation "norm". For teacher educators, this study suggested the necessity of attending to continuous classroom discourse and the long term consequence of interactions. Considering classroom interactions as the process of constructing learning related, shared mediated resources may help teachers and teacher educators to reflect on teaching practices and understand teaching effectiveness from a more constructive, long-term perspective. While activity theory has established theoretical account of how activity system can evolve locally on itself, little work has been done to illustrate such process with data. Through thick description on the construction and functioning of shared mediational resources in science classroom, this study also provided a concrete account corresponding with such theoretical processes.

## Reference:

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22,* 797–817.

Berland, L., & Reiser, B. (2009). Making sense of argumentation and explanation. *Science Education, 93*(1), 26-55.

Crawford, T., Kelley, G.J., & Brown, C. (2000). Ways of knowing beyond facts and laws of science: An ethnographic investigation of student engagement in scientific practices. *Journal of Research in Science Teaching, 37(3),* 237-258.

Cazden, C. (1988). *Classroom discourse: The language of teaching and learning.* Portsmouth, NH: Heinnemann.

Dawson, V. & Venville, G. (2009). High-school students' informal reasoning and argumentation about biotechnology: An indicator of scientific literacy? *International Journal of Science Education, 31*, 1421-1445.

diSessa, A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Hillsdale, NJ: Erlbaum.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*, 287– 312.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science.* Buckingham: Open University Press.

Duranti, A., & Goodwin, C. (Eds.). (1992). *Rethinking context: Language as an interactive phenomenon.* New York: Cambridge University Press.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education, 38,* 39 – 72.

Engestrom, Y., (1987). *Learning by expanding*. Orienta-konsultit, Helsinki.

Engeström, Y., Miettinen, R. & Punamäki, R.L. (1999). *Perspective on activity theory*. Cambridge, U.K.: University Press.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20(4),* 399-483.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education, 88,* 915–933.

Erickson, F. (2004). *Talk and social theory: Ecologies of speaking and listening in everyday life*, Cambridge, Polity Press.

Fuller, S. (1997). *Science.* Buckingham, UK: Open University Press.

Gee, J. B. (1999). *An introduction to discourse analysis: Theory and method.* London, England: Routledge.

Geertz, C. (1973). Thick description. In C. Geertz (Ed.), *The interpretation of cultures* (pp. 3–30). New York: Basic Books.

Gumperz, J.J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.

Hammer, D. (2000). Student resources for learning introductory physics. American Journal of Physics, *Physics Education Supplement, 68*, S52-S59.

Jim´enez-Aleixandre, M.-P., Rodr´ıguez, A. B., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education, 84,* 757 – 792.

Kelly, G. J. (2005). Discourse, description, and science education. In R. K. Yerrick & W. M. Roth (Eds.), *Establishing scientific classroom discourse communities: Multiple voices of teaching and learning research* (pp. 79-104). Mahwah, NJ: Lawrence Erlbaum Associates.

Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education, 20,* 849 – 871.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77,* 319–337.

Kuhn, D. (1991). *Skills of argument.* Cambridge, MA: Cambridge University Press.

Latour, B., Woolgar, S. (1986). *Laboratory life: The construction of scientific facts.* Thousand Oaks, CA: SAGE.

Lemke, J.L. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.

Louca, L., Elby, A., Hammer, D.,&Kagey, T. (2004). Epistemological resources: Applying a new epistemological framework to science instruction. Educational Psychologist, 39(1), 57 – 68.

McNeill, K., Lizotte, D., Krajcik, J., & Marx, R. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, 15(2), 153-191.

National Research Council (NRC). (2000). *Inquiry and the National Science Education Standards*. Washington, DC: National Academy Press.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.

Toulmin, S. (1958). *The uses of argument.* Cambridge: Cambridge University Press.

van Zee, E.H., & Minstrell, J. (1997). Using questioning to guide student thinking. *The Journal of the Learning Sciences, 6, 229-271.*

Vygotsky, L.S. (1978) *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. Journal of Research in Science Teaching, 39(1), 35 – 62.

# Knowledge Transmission and Engineering Teaching

Amy Zhang and Monica Cardella, Purdue University, 701 West Stadium Avenue, West Lafayette, IN, 47906, USA
Email: zhang212@purdue.edu, cardella@purdue.edu

**Abstract:** This paper explores the concept of knowledge as well as the process of knowledge transmission, both in general and specifically regarding engineering knowledge. This study focuses on twenty-four parents who discussed the activities that they engage in to help their children learn about engineering during hour-long interviews. The intergenerational transmission of knowledge that is described in the interviews is analyzed through 1) social structural and cultural analysis, and 2) a case study. We find that parents teach children what they perceived as engineering knowledge, which is not necessarily true but rather subjective, and believed to be a fact based on the parent's other forms of knowledge learned in the past. Understanding this will help us look at engineering education from a new perspective, and the application of this understanding could be used to help provide non-engineering parents alternative ways to support their wish to explore engineering knowledge with their children.

## Introduction

This paper deals with pre-college students' engineering education within the backdrop of the current state of pre-college engineering education in the United States: while engineering has traditionally not be explicitly taught in the K-12 schools as a subject, there is growing interest in incorporating engineering in K-12 schools (e.g. NAE and NRC 2009). A small number of states have even begun adopting engineering curriculum standards. However, as this is a recent change to K-12 education, there is little research on engineering pedagogical content knowledge or the impact of engineering instruction on students' understanding of engineering. As such, rather than focus on expert teaching practices exhibited by K-12 teachers, this study focuses on the teaching practices of engineering parents. While this expertise may not always translate to practices that classroom teachers can adopt, we believe that some of the parents' practices can inform classroom instruction as well as the development of resources for non-engineer parents. Additional motivation for the study is based in a desire to increase diversity in engineering at the college-level and beyond (i.e. post baccalaureate studies and industry practice).

In order to investigate the practices that parents with engineering backgrounds employ to help their children learn about engineering, we interviewed 24 parents who self-identified as parents of children aged 2-18 who "help their children learn about engineering." As we selected our study participants, we were careful to include parents from different engineering backgrounds in addition to academic vs. industry vs. mix-of-industry-and-academic backgrounds, as well as parents of children of difference ages and sexes. Each parent participated in a semi-structured interview that last approximately 60 minutes. We began by asking about their children; their children's schools; their academic background. Then we asked them about how they helped their children to learn about engineering, and asked follow-up questions based on their responses.

The interviews were conducted among 24 participants, among whom there are 12 males, and 12 females. 14 parents out of 24 are faculty members, 2 are PhD students, and 8 are in industry. Both of the parents who are PhD students have had respectively 9 to 30 years working experience in industry prior to the PhD program. The parents' work and educational backgrounds touch on 20 specific engineering fields, such as civil engineering, biomedical engineering, aerospace, electrical engineering, and industrial engineering. Out of their 50 children, 26 are girls and 24 are boys. Their ages range from 17 months to 29 years old. Besides 11 outliers, there are 39 children who age from 2 to 18. The interview questions have covered parents' background, interactions with children (content they teach their children, teaching strategies, and children's reactions), parenting ideology, and parents' own understandings of engineering.

In this paper, we use these interviews to explore parents' roles as teachers; how knowledge is transmitted intergenerationally; and interplays of socio-cultural factors and parents' own history of learning that impacted parents' activities. We do this by focusing on four participants: Laura, an Asian mother of one son and one daughter and is pursuing a PhD in software engineering; Liz, a Professor teaching biomedical engineering who has a four year old son; Tom, a father working in the industry with an engineering background covering several fields (metallurgical engineering, material engineering, and process engineering) and who has a 2.5-year-old daughter and a 5.5-year-old son; and Aaron who is a father of a 4.5-year-old son and is a professor with an electrical engineering background.

We begin this paper by discussing the concept of knowledge. By dividing knowledge into specific knowledge and common sense knowledge, we are able to analyze interplays of the two during the process of the intergenerational knowledge transmission– how the former is based on and restrained by the latter which is deeply embedded in and in turn reinforces the former. The process of intergenerational knowledge transmission is analyzed on two levels – social structural and cultural analysis and processual analysis. In order to understand the meaning of parents' teaching activities, we have to look at the social structure and cultural environments in which these practices are performed. Structural analysis as a methodology thus helps us understand 1) that parents' educational practices are shaped by the basic principles of their common sense knowledge, 2) their ideology of education functions as to ensure the common sense knowledge to be put in their daily practices and obeyed, 3) how different social systems such as families, engineers, and different ethnic groups are integrated, 4) how individual parents justify their practices through scrutiny, and 5) how through the interaction between individual parents and the society, both the parents' goals and the social expectations become realized.

## Knowledge transmission through teaching and learning

In a footnote of Foucult's Archaeology of Knowledge (Foucult, 2002:16), his translator A. M. Sheridan Smith wrote:

> "The English word 'knowledge' translates the French 'connaissance' and 'savoir'. 'Connaissance' refers to a particular corpus of knowledge, a particular discipline – biology or economics, for example. 'savoir' which is usually defined as knowledge in general, the totality of connaissance is used by Foucult in an underlying, rather than an overall, way. He has himself offered the following comment on his usage of the terms:
>
> > 'By connaissance I mean the relation of the subject to the object and the formal rules that govern it. Savior refers to the conditions that are necessary in a particular period for this or that type of object to be given to connaissance and for this and that enunciation to be formulated'
>
> Throughout this translation I have used the English word, flowed, where the meaning required it, the the appropriate French word in parentheses."

There are two types of knowledge implied in the above message which are not mutually exclusive. One is expertise or skills gained in a specific field through education or learning; the other is implied above by Foucult, that is – different epistemologies we believe in and practice every day to know and make sense of the world (i.e, common sense knowledge). However scientific the former might be, it is always limited by the latter which we practice almost unconsciously in everyday life. However encompassing our common sense knowledge might seem to be, our specific knowledge gained from systematic learning in turn reconstructs our common sense knowledge by a series of rejections, confirmations, and readjustments. Therefore, as Hansen states, "the transmission of knowledge is subject both to conservative forces and to tendencies toward continual redefinition" (Hansen, 1982:26).

In our case, the specific knowledge being transmitted from parents to children is more transparent. The knowledge is everything that parents think is engineering related – fixing a household item, building an electric circuit, practicing mathematics skills, understanding how things work, etc. The common sense knowledge, however, is more complicated, pervasive, and taken for granted during most of our waking hours.

According to Hansen (1982), common sense knowledge – which he calls cultural knowledge – is a set of maxim, ideas about human nature, aesthetic preferences, values, affective patterns, beliefs, etc. (Hansen, 1982:25). According to Leiter (1980) and Schutz, common sense knowledge does not only include the "rules of thumbs that are vague, contradictory, and self serving", but it can be studied from three dimensions – the stock of knowledge, the natural attitude of everyday life, and the practices of common sense reasoning (Leiter, 1980). Much of these dimensions hinges upon the various assumptions people make about each other and each other's reasoning. In this study of parents' teaching experiences, we examine the process of knowledge transmission by looking at the interactions of these two types of knowledge and how parents reorganize these types of contradictory as well as mutually sustaining knowledge to selectively teach it to the next generation.

## Structural & Cultural Analysis

Knowledge transmission reflects social structure. Social structure is viewed by Sewell (1992) as "dual", as "both the medium and the outcome of the practices which constitute social systems". Social structure also "differs in 'depth' (how pervasive, invisible, and taken-for-granted their schemas are), and 'power' (how great the resources they generate from)" (Sewell, 1992:22-6). Practices are therefore enabled as well as constrained by social structure, and these practices in turn make the transformation as well as the continuity of social structure possible. Parents' everyday teaching practices such as the types of knowledge selected by parents, their children's expected responsibilities and privileges, and the descriptions of children's "good" and "dissatisfying" performances (which reflects their core values) are enabled and constrained by what facilities are around the parents, what methods are favored by the mainstream culture (e.g., what they see from the mass media), what parents can afford to do, what parents' peer groups are doing, and how parents cope with their children's school curriculum.

In our 24 interviews, there are two channels through which parents teach their children engineering related knowledge: material resources and daily interactions. These two channels are not necessarily mutually exclusive. The most commonly used material resources are manipulative toys, computer programs, websites, books (either literature or science oriented), TV programs and DVDs, and trips to museums or exhibitions. The use of these resources is largely determined by the manufacturing and consumption preferences of the society, distribution of knowledge and skills, relationship between specific knowledge and its financial reward as well as its social reward, access to the resources due to the parents' social class, parents' financial ability, household locality, and the children's age, gender, birth order, and their assumed personality.

For example, the parents' financial ability decides what types of toys are on the top of their hierarchy and for what purpose the parents are buying the toys. Laura (1), one of our parents who participated in the study, mentioned that she used Barbie dolls as tools to help her daughter practice counting when she was 2 years old. In contrast, imagine a low-income family with many children. Here, the parents might use objects that are more practical and affordable or objects that they already have in their possession. The purpose of buying a Barbie doll is different, too. For a low-income family, parents may buy their children a Barbie doll based on their wish to provide their children a toy that other children possess. In contrast, Laura might have multiple different considerations. For example, a different Barbie doll would increase the variety of toys for her children, and the toy itself can be used as a teaching mechanism (and Barbie's latest career as software engineer, though announced after our study concluded, may provide yet another opportunity for parents to teach their children engineering concepts) .

Another channel through which parents teach their children engineering concepts is daily interactions. Some parents take 10 to 20 minutes every night to answer some questions before their child goes to bed, usually curious questions about nature. In doing so, they claim to encourage their children's curiosity, a characteristic which is highly valued in a middle class American family with both parents systematically and well educated. Some parents invented a mini project such as building an electric circuit or fixing an old computer to engage their children in the problem solving process in a "natural" environment. Most parents have taken their children to work, and some of them have exposed their children to specific objects they use at work. According to the Neo-Marxist conflict theory (Collins, 1985), such a way to organize interactions reflects the social status group of the parents. Additionally, through associations with members in the same group, the parents share common status cultures such as language styles, parenting styles, requirements for education, interpersonal dynamics, values, topics, etc (Collins, 1985). Such ways in which educational interactions take place make it possible for the children to become technologically capable, curious, creative, and able to fix problems. These are expectations that overlap in both the macro American culture and the engineering culture specifically and are constructed under these certain social structures. In meeting the expectations, the social structure in turn is reinforced.

The above analysis is not to imply that these parents – who share a similar social class status, academic backgrounds, parenting styles, and expectations for children – as a group are homogenous. The knowledge these parents choose to pass on to their children through selected strategies reflect the integrated core values of 1) the mainstream Anglo-American culture, 2) the microscopic engineering culture, 3) various subcultures (race, gender, nationality, etc.), and 4) the parents' historical experiences. For example, students in American culture are encouraged to take initiative and be motivated to learn. "Students are expected actively ask questions and participate in class discussions and other activities; they are frequently rewarded for contributing to the class and giving critical and constructive ideas" (Pai & Adler, 2001:221). Self-motivation, independence, curiosity, and creativeness are considered as desirable qualities (Pai & Adler, 2001). Children are expected to reflect such qualities as well. Liz, who is a professor teaching biomedical engineering, complained that her son doesn't have initiative to learn. When we asked her to describe her son, she carefully said that her son is "different" from other kids in the sense that he doesn't seem to have a lot curiosity, and he doesn't ask a lot questions. Liz's worry came from the awareness that her son doesn't have some of those core values that are preferred by this society: curiosity and initiative. While every other kid is praised for being curious, Liz is concerned that her sons' passive learning characteristic will not help him succeed.

As Swidler (2001) postulated, social structure depends on the mutual reproduction of schemas and resources. Schemas are the semiotic codes shared by a group of people and used by them to make sense of the world. A schema is what makes a resource meaningful as a resource (2001: 78). Take our parent study as an example. These parents with engineering backgrounds are well educated by formal institutions with at least a Masters degree while some have a Ph.D. The engineering schemas – such as solving an engineering problem –

that are deeply embedded in heavily scripted engineering classroom interactions are internalized. Thus when parents who share the same engineering schemas see a child taking a toy apart, they think of the action as initiating a problem which leads to the next step – putting the toy back together. However, parents who don't share the same schemas are more likely to blame their children for breaking the toy and possibly causing unnecessary financial repercussion. The schemas parents internalized during their own learning process became their own rationale they use to interpret resources and interact with their children who in turn learn to interpret things by using the same schemas. With the ongoing mutual reproduction of resources and schemas the social structure is sustained.

Different parenting styles may also be attributed to different sub-cultural backgrounds. Aaron, who is a professor described how he took his son down to their basement and explained to him the structure and functions of the water pipes. Laura, who is a Chinese woman pursuing her Ph.D. in America says that she is not very satisfied with the mathematics education here. "There's too much play-work here in American elementary schools", she said, "I still think the Chinese way is better". She gave her children math practices to do and graded them like a teacher. Our interviews suggest that parents within engineering fields but with different specializations have different understandings of what engineering is and therefore have different focuses on teaching what they think is engineering related. Parents who see engineering as problem solving are more inclined to initiate a small project to get their children involved, and the focus is on how things work. One parent who calls herself a "soft engineer" (which is generally associated with problem solving, design, communication skills, and teamwork) focuses more on the basic scientific and numeral skills which she thinks are fundamental and crucial to engineering. Tom, who has worked in industry for a long period of time (as opposed to other parents who are in academia) sees that engineering is not always problem solving but most of the times is just maintaining, to "keep the machine running" as he said. However, his teaching strategies don't seem to be too different.

According to Durkheim's social order theory, the study of social structure focuses on the moral order, the central value system that though created by people, has an independent and external existence and acts as a constraining and conditioning force upon individual people (Meighan and Sirah-Blatchford's, 2003: 252). Parsons, like Durkheim, saw social order achieved through the operation of an integrating system common to all members of society; yet he also emphasizes how individuals constantly adjust themselves through scrutinizing the process of socialization (2003:256). Structural analysis as a methodology thus helps us understand 1) that parents' educational practices are shaped by the basic principles of their common sense knowledge, 2) their ideology of education functions as to ensure the common sense knowledge to be put in their daily practices and obeyed, 3) how different social systems such as families, engineers, and different ethnic groups are integrated, 4) how individual parents justify their practices through scrutiny, and 5) how through the interaction between individual parents and the society, both the parents' goals and the social expectations become realized. Further, the integration of structural analysis, cultural analysis, and Swindler's understanding of structure as the mutual reproduction of schemas and resources provides us a different perspective that gives meaning to parents' teaching activities. Therefore, we not only focus on the relationship between individuals and society, but also the meaning of individual practices under certain schemas constructed by the members of the society (e.g. a child taking a toy apart is perceived by engineering parents as an aspect of curiosity and an initiative step to problem fixing). However, the study of social structure and culture focuses on the relatively stable features of behavior and context and the patterned arrangement of relationships among individuals and groups while leaving the problem of process untouched. We are also interested in the process of how these schemas came into being and how parents reorganize the messy and conflicting knowledge from different levels of social systems and transmit the knowledge to the next generation. In the following case study, we delve into these processes by analyzing a specific interview with a female engineer who we think is particular in representing both conflicts and integration of all kinds of knowledge she is trying to pass to her children.

## Case Study

Our case study will center on an interview with a female engineer and parent. In order to better reveal the process of intergenerational knowledge transmission, in which the parent tried to put different pieces of knowledge (engineering knowledge, the parent's own cultural background, and the mainstream American culture perceived to be true by her) together, we follow Hansen's suggestion (Hansen, 1982, 1990) by dividing the following analysis into several steps. These steps will allow us to see how the social structures are created, maintained, challenged, and modified over time as well as process an individual's adaptation necessary to his/her changing environment (Hansen, 1990: 192). The steps are: 1) the definition of the situation; 2) the cultural significance associated with communication channels used to encode and decode communications; 3) the interplay among channels of transmission; 4) the communicative-interpretive repertoires of participants, including communicative competence in the codes being used, 5) the communicative strategies participants used to realize their respective interests and purpose, and 6) the role and identity attributions.

Laura is a Chinese woman pursuing her Ph.D. degree in engineering, specialized in software programming, computer networks, and information security. Her husband is also an engineer and currently working in industry. Prior to pursuing her Ph.D., Laura had eight years of industry experience. Laura has two children, a son and a daughter. In the following conversation, you will see Laura trying to tell the interviewer what kind of engineering related knowledge she taught her children. She also stated that she was a "soft engineer" and that many things she had to do at work are math related. Therefore, she viewed math as fundamental for engineering.

> Interviewer: so you started doing that after they have learned some math at school?
>
> Laura: yes . but actually , their math, uh-mostly, I uh taught them math because their math is more advanced than what they are learning at school?[ Interviewer: uh-hmm] so I uh, the-they learned their math at home, basically. ((chuckle))
>
> Interviewer: Oh, OK. So-
>
> Laura: [Schools, especially middle schools here are not too demanding here in America? I would say American education here, eh- even in the best middle schools, in west Lafayette, I know a lot of parents, the way they teach math, I think it's not demanding enough, so that if I don't teach them at home, I feel like their talent will get wasted. So ((chuckle)), that's how I feel.
>
> Interviewer: So are there other ways in which you tried to explain math to them?
>
> Laura: um, see, uh-here in America, we like to say let's play games and do math. [Interviewer: uh-hmm]And we don't do it that way. I do it in the Chinese way. I came from china, and I, and I-learned my way there. Ah-I teach them Chinese way((laugh))
>
> Interviewer: [OK. So it's kinda like complementary-
>
> Laura: [yeah I thought it was very important. ((laugh)) [A: OK. That's interesting.] I actually think the Chinese way of teaching math is better? ((laugh))

According to her description, her definition of the current situation was that she is not satisfied with both the depth of the math education in America in general and the play-work teaching strategies she has observed from other parents and schools. The school was wasting her children's talent. She expected that teaching her children math the Chinese way would stop her children from falling behind where they were supposed to be, were they raised in China.

Apparently there are two conflicting communicative channels that are associated with two different cultures. One channel is associated with mainstream American culture and stresses the importance of autonomous study, independence, initiative, and having fun, especially at the level of elementary school. The other channel through which the knowledge is transmitted is associated with Chinese culture which values cooperative study, passiveness (e.g. students are expected to follow the guidance and authority, and learn through watching, listening and emulating), memorizing the knowledge, and heavy practices. In general, American school personnel see schooling as a process of developing the whole person (Pai &Adler, 2001:220-223). Thus non-academic activities are considered an important part of schooling. Playing, for example, is perceived as a strategy to initiate children's curiosity to learn as well as a way to develop the children's social skills. As Spindler and Spindler (1990:37) generalized, one of the core values of the Anglo-American society is a sociable, get-along-well orientation. The ideology of early children's education is also to explore and support their own interests instead of pushing them to any one direction (which is almost a taboo in the interviews with parents). However, in a Chinese mom's eyes, playing at school – where only academic related activities are supposed to take place – can be a distraction to the child's intellectual training and cause anxiety for the mother. Being a Chinese mom, Laura also expects her children to be among the top students in their classes. Even though she claimed that she didn't intend to push her children, she was very proud when she told me that her son had won several math competitions and her daughter – who didn't show very much enthusiasm in math – was a year ahead of her classmates. Facing this anxiety, Laura felt the impulse to change the situation. She then went on to tell me how she achieved this change.

Another channel through which Laura teaches her children engineering related knowledge is associated with the history of her engineering learning process. Most of Laura's education was done in China where she also had 8 years of experience working in engineering industrial practice. She is also currently studying for her Ph.D. in engineering in an American institution from which she has internalized the core engineering values in the U.S. through her learning practices. Her perspective of what engineering is can be very different from the mainstream perspectives as we shall see below:

> :: (My daughter)-she's very self sufficient,? She, she, she cooks her own pancakes,?((chuckle)) [A:oh] so we think that would related to engineering. you wouldn't think cooking and engineering (*are related*), but engineering, uh, you follow a direction, mix the eggs and flour together, ((laugh))[A: right]. I think that'll will be good-]
>
> Interviewer:
> [It's a, i-it's a process of production?
>
> Laura: Right. Production. And also you know the modern grinder, the mixer. So 'cause grandma has to do it by hands, she does it by hand. And with the mixer, they know this is the-because there is engineer.(ing).

The above conversation shows that engineering is perceived by Laura as a symbol with two meanings: 1) engineering is to "follow a direction" or follow instructions, and 2) engineering is to design innovations which improve our lives. The second meaning is a commonly shared meaning within the American engineering culture, whereas the first one is gained from Laura's history of learning and working in China. While the second meaning is articulated frequently and almost throughout the entire interview, it is the first meaning of engineering – follow the instruction that Laura uses to construct her real parenting practices. For example, feeling the anxiety that the school is not demanding enough at math teaching, Laura brought home many math practice sheets for her children to work on, and "I grade them like a teacher", she said. If the children did the problems wrong, she would circle the wrong answers and ask them to redo the questions. Laura also bought her children the Stanford EPGY (Educational Program for Gifted Youth), a computer based program for math, and asked them to practice by following the program's instructions at least half an hour each day. Both solutions consist of two types of instruction following: one is to follow Laura's own instruction to finish the work she gave to her children, the other one is to study math by following the instruction of the program.

When asked about what she usually taught to her children, just like all of the other parents we have interviewed, Laura felt the necessity to immediately claim that she didn't really teach her children anything specific because she wanted her children to be free to choose what they like to do. However, in the later description of how her daughter learned math, she first showed some dissatisfaction since her daughter wasn't as motivated to learn math as her son; she then said she felt lucky because her daughter doesn't really hate math even though she didn't show much enthusiasm; at last, she was proud to say that her daughter, although not as successful as her son, was still doing very good at math and her level was one year more advanced than her peers. Therefore, by articulating what is expected by mainstream society and then actually practicing under completely opposite schemas, the social values are sustained and Laura's personal purpose is realized. Laura also did not have to be fully conscious during the process of readjustment.

As we have shown above, the interview with Laura was full of contradictories. As a self-selected participant, she was attracted by our flyer looking for parents who transmit engineering knowledge to their next generation and told us that this project was interesting. She might have been interested in our project for two possible reasons: 1) our requirement of participants confirmed her self-presentation with multiple identities – a mom, an engineer, and a responsible mom (for she gives "good" education to her children), and 2) she hopes our research will provide more learning opportunities. Either reason implies her will to expose her children to engineering knowledge. However, she came to us and claimed that she had never intentionally taught her children anything specifically about engineering like almost every other participant does. She tried very hard to leave us an impression that she is a good mom, and therefore will give her children enough freedom to explore what their interests are and be supportive. However, this concept of a good mom is challenged when she saw her children doing activities that are non-academic and thought such activities might hinder her children's potential to become successful, for being a good mom also means having successful children. It is possible that she did not teach her children engineering knowledge specifically as she stated, yet the fact that she was attracted to our research was a self-presentation as an engineer who is self reflexive and aware of the many benefits our society has credited to the engineering field.

As shown above, when unfolding the process of knowledge transmission it is important to understand the definition of the situation (what is happening, who is present, and what is expected to happen next), the interactions among different channels through which the knowledge is transmitted, how individuals either consciously or sub-consciously manipulate the meanings associated with different social groups, and the process of role making within such situations. It is in people's interactions that we find things meaningful.

## Summary

In our 24 interviews, parents discussed teaching their children engineering related knowledge: through two primary channels: material resources and daily interactions. The most commonly used material resources are manipulative toys, computer programs, websites, books (either literature or science oriented), TV programs and DVDs, and trips to museums or exhibitions. The use of these resources is largely determined by the manufacturing and consumption preferences of the society, distribution of knowledge and skills, relationship between specific knowledge and its financial reward as well as its social reward, access to the resources due to the parents' social class, parents' financial ability, household locality, and the children's age, gender, birth order, and their assumed personality. The daily interactions through which parents claimed to teach their children engineering knowledge have revealed different forms of knowledge that have been transmitted along with engineering knowledge. During the analysis of the process of knowledge transmission, we were able to see both conflicts and integrations of different forms of knowledge, and how parents tried to put the pieces together in order to achieve their goals of education. Both channels showed that knowledge is messy and exists in many different forms, and that when engineering is taught by parents to their children as one type of knowledge, other forms of knowledge are also inevitably active in this process. These forms of knowledge (culture, specific knowledge, schema, practice, and etc) help in shaping and in turn reinforcing each other.

Understanding the concept of knowledge and the process of knowledge transmission allows us to come out of our own bubble and look at engineering education from a different perspective. During the learning process in general, we draw the information from our stock of knowledge, assume the existence of order and pattern, and seek to organize our perceptions in such way to discern it (Hansen, 1990:4). We categorize the information by using our internalized hypotheses learned from the past and orient our ways to export the knowledge by reference to these hypotheses. During the engineering education process within the household in particular, parents teach children what they perceived as engineering knowledge, which is not necessarily true but rather subjective, and believed to be a fact based on the parent's other forms of knowledge learned in the past. In the future study, we intend to apply our understanding of knowledge transmission to k-12, and pre-school engineering education. We intend to build on the study presented in this paper through further interviews with parents as well as observations of parents and children at play. We also plan to extend the work by providing resources for non-engineer parents to use as well as using our research findings to develop activities to use in formal (i.e. classroom) settings. The study described in this paper also shows us that increasing access to engineering education for children who might not traditionally be exposed to engineering is not as simple as providing ideas for activities that parents or teachers can engage in with children; instead the parents' and teachers' prior engineering experiences and cultural experiences can deeply impact the way that they approach engineering education. Thus these engineering and cultural beliefs must be understood in order to impact pre-college engineering education.

## Endnote

(1) All names have been replaced with pseudonyms.

## References

Cole, M., Gay, J., Glick, J. A., and Sharp, D. W., (1971). The cultural context of learning and thinking. An exploration in experimental anthropology. London, Methuen & Co Ltd.

Collins, R. (1985). Functional and conflict theories of educational stratification. Schools and society: A reader in education and schooling. Palo Alto, CA: Mayfield, In J. H. Ballantine.

deMarrais, K. B. & LeCompte, M. D. (1999). The way schools work. A sociolological analysis of education. New York, Longman.

Dupuis, A.M. & Gordon, R. L. (1997). Philosophy of education in historical perspective. Second edition. Lanham, University Press of America, Inc.

Foucault, M. (2002). The archaeology of knowledge. Translated by A. M. Sheridan Smith. Park Square, Routledge Classics.

Hansen, J. F. (1990). Sociocultural perspectives on human learning. Prospect Heights, Illinois, Waveland Press, Inc.

(1982). "From background to foreground: toward an anthropology of learning." Anthropology &Education Quarterly 12(2): 13.

Kunda, G. (1922). Engineering culture. Control and commitment in a high-tech corporation. Philadelphia, Temple University Press.

Leiter, K. (1980). A Primer in Ethnomethodology. New York, Oxford University Press.

Lincoln, Y. S., Guba, E. G., (1985), Naturalistic inquiry, Sage Publications, New Haven, CT.

Meighan Roland, S.-B. I. (2003). A Sociology of educating. London and New York, Continuum.

Montgomery, J. D. (1997), Values in education. Social capital formation in Asia and the Pacific. Edited. Hollis, Hollis Publishing Company.

Moore, R. (2007). Sociology of Knowledge and Education. New York, Continuum International Publishing Group.

(NAE and NRC) National Academy of Engineering and National Research Council (2009). Engineering in K-12 Education: Understanding the Status and Improving the Prospects. Washington, DC, National Academies Press.

Pai, Y., and Adler, Susan A. (2001). Cultural foundations of education. Upper Saddle River, New Jersey, Columbus, Ohio, Merrill Prentice Hall.

Peak, L., (1991), Learning to go to school in Japan, University of California Press, Ypsilanti, MI, .

Spindler, G., Spindler, L., (1987), Interpretive Ethnography of Education: At Home and Abroad, Lawrence Erlbaum, Prospect Heights, IL.

(1992), Cultural process and ethnography: an anthropological perspective, in M. Lecompte, W. Milroy, J. Preissle (Eds), Academic Press, Hillsdale, NJ.

Sewell, W. F, (1992). "A theory of structure: duality, agency, and transformation". The American Journal of Sociology, Volume 98, No. 1 (Jul., 1992), 1-29.

Swiddler, A. (2001). What anchors cultural practices. The Practice Turn in Contemporary Theory. T. R. Schataki, Cetina Karin Knorr, and Savigny Eike von. London and New York, Routledge.

# Listen to each other: How the building of norms in an elementary science classroom fosters participation and argumentation

Suna Ryu, William A. Sandoval
University of California, Los Angeles, Los Angeles, CA 90095
Email: sunaryu@ucla.edu, sandoval@gseis.ucla.edu

**Abstract:** The present study explores and documents the process of building social and argumentation norms in a combined third- and fourth-grade science classroom and examines how the processes regulate students' participation patterns, including the students' roles. This study also investigates the positive potential of negotiating norms and taking on intellectual participatory roles for fostering scientific argumentation. The analysis focuses on how one norm, *listening to each other*, contributes to distributing participation rates and forces students to take more critical audience and intellectual roles. Students also show improvement in argumentation in terms of frequency and epistemic quality.

## Introduction and Background

Argumentation is one of the central discursive practices of science, and has been highlighted as an essential part of science instruction (Driver, Newton, & Osborne, 2000; NRC, 2007). This emphasis stems from recently articulated views of science that underscore the deeply social aspects of knowledge construction in science. Due to this growing importance, a body of research has developed about student argumentation and how to support it(see Eduran & Jimenez-Alexaindre, 2008). In this research, scholars do not always make clear what they mean by argument or argumentation. We use the argument here to refer to any "text" (written or oral) that involves one or more claims and justifications for those claims. Argumentation is thus the process by which such texts get produced. We also view learning to argue as an important goal in itself for science education, as Driver et al. argue, as opposed to arguing as a particular method of learning concepts (Andreissen, 2006).

From this body of research scholars agree that learning to argue well requires more than just acquiring cognitive skills. Argumentation, by definition, is a social practice (van Eemeren et al.,1996). Consequently, student argumentation has a complex relationship with other social aspects of the classroom culture. Hence, learning to argue well involves the mastery of a complex discursive practice governed by a set of norms that has been refined over a long period of time. Fostering argumentation in the classroom community requires a fundamental shift in classroom culture, in particular, an innovative change in discourse use (Duschl, 2008).

Among the issues in the typical classroom culture, teacher-dominated IRE types of discourse and the vertical relationships between teachers and students have a long history of being nominated as fundamental issues that hinder students from engaging in active dialogical argumentation (Driver et al, 2000; Dushl, 2008) Recent research also reports that such discourse patterns and vertical power relationships are often mirrored when students interact with each other. That is, academically high-achieving students tend to take on the teacher's role as the evaluator/answerer/clarifier in small groups and reproduce IRE types of discourse (Webb et al, 2006). For instance, when a group member asks for help, smart students simply give the answers to the student. Hence, the reproduced discourse among students may prevent students from voicing different opinions or arguments. Research also indicates that vertical power relationships among students result in unequally distributed participation rates. A stratified pattern of participation tends to be found: some students often dominate the classroom conversation (Cornelius, 2008).

To overcome the problems of traditional classroom discourse, some researchers point out that a teacher's adoption of a new discourse can be the key to improving classroom discourse (Driver et al., 1994, 2000; Erduran, Simon, & Osborne, 2004; Osborne, Erduran, & Simon, 2004). In addition, many studies across disciplinary areas have explored participant structures in which teacher and students re-define their roles to improve their discourse. For example, studies have documented how teachers can newly position themselves as a facilitator, and/or negotiator, rather than as an evaluator, or answer the student so that students take the main role as knowledge co-constructors. Brown and her colleagues (Brown & Campione, 1990; Brown, 1992; Brown et al., 1993) facilitated the engagement of students by encouraging them to take specific roles in a jigsaw, and the students generated, manipulated, and discussed their ideas in a public forum. Herrenkohl and Guerra (1998) highlighted three practices, including monitoring one's own comprehension of another's ideas, coordinating theories with existing evidence, and challenging the claims put forth by others. They found that students were more likely to coordinate claims and evidence by taking an audience role that encouraged the students to provide clarifying and monitoring comments associating with taking intellectual roles. The researchers argued that students can participate in a social practice common to scientists by having students engage in episodes that request coordination between theories and evidence. In the study, an important aspect of dialogical argumentation—ideas and perspectives are exchanged, refined, and evolved—was highlighted by asking students to take the role of an audience that is actively involved in initiating engagement rather than simply taking on the role of passive intellectuals. Van Zee and Minstrell (1997) found that the teacher's use of the "reflective toss," pushed students to elaborate on each others' and their own ideas, and become more active participants in whole-class discussion. These studies illustrate the kind of classroom community and culture may promote students' argumentation. As Roseberry et al. (1992) argued, students not only take on new roles and more actively participate in classroom discussion, but they also are more likely to provide better scientific explanations that relate evidence to claims.

However, many issues about how a teacher's discursive practice with students (the establishment of specific discourse norms that can help students to reorganize and regulate their roles, and how these things may work together) can foster students’ argumentation remain unexplored. We don’t really know how teachers create these classroom communities. In addition, rarely studied is how the established norms and roles, guided by the teacher may influence students' power dynamics. In this context, I address the following three questions.

1) When a teacher initiates social norms in the context of the elementary science classroom, what participation and argumentation norms seem to emerge and become negotiated?
2) How does the process of initiation and negotiating norms help students to reorganize/regulate their roles in small group discussion?
3) How do the negotiation of norms and the taking on of participatory roles foster scientific argumentation?

### Setting and Participants

This study took place in a combined third- and fourth-grade classroom in a laboratory elementary school at a large public university in the western United States. The population of the school approximately mirrors its state's population in ethnicity and socioeconomic class. The class included 21students: 12 third-graders (6 boys and 6 girls) and 9 fourth-graders (5 boys and 4 girls). Ethnically, there are 18 Latino students, 2 Caucasian students, and 1African-American student because this class is part of a dual-language (English and Spanish) immersion program at the school. There were three one hour periods of science each week. For a unit consisting of four or five weeks, students began by creating big ideas and questions. In this science class, students were asked to research science topics, often by reading books and websites. The students also conducted investigations and collected data. For example, students investigated the characteristics of magnetic force by putting magnets into the sand box on the playground of the school, by observing the number of washers required to break two magnets, and by drawing graphs and comics about the magnets. The students were then asked to represent their findings and present them to other students. As they worked, students were encouraged to participate in discussions by presenting one another’s opinions and by resolving disagreements.

Students spent 65% of their time, on average, in small groups, rather than together as an entire class. In a small group, students are assigned to one of various roles, such as a starter (who explains the group’s goal and plan), recorder (who records the data and presents them to the class), helper (who asks for helps from the teacher if the students have questions or need other help), or a gatherer (who prepares the materials for the experiments). Once a week, the teacher assigns students new roles. The teacher, Ms. Green, had been teaching 38 years at the time of the study. Ms. Green originally majored in English, but she had more than 20 years of experience teaching science at the elementary school level.

## Data Sources and Methods

A qualitative case study design was used in this study. A case study design was employed because it provides an in-depth description of a particular activity and situation (Creswell, 2003; Meriam, 1994; Miller & Salkind, 2002) that verifies how the norm is developed in a classroom (natural context) and examines students’ roles and power dynamics.

I used a combination of classroom observation (field notes and video data), teacher interviews, and student interviews to answer my research questions. Videotaping and taking field notes were essential because this study requires fine-grained analyses of student discourse use and interaction. The video record enabled in-depth micro-analysis of interaction (Erickson, 1992). Brief field notes were also recorded in order to capture classroom contexts that might occur outside the camera’s range. The video data for this paper include 18 science lessons (approximately 20 hours) from the first six weeks of the school year. The video data were first logged, and episodes determined to be of analytic interest were fully transcribed. The transcription included partcipants' speech, gestures, facial expressions, and behaviors. To analyze the data, a text management program, Atlasti, was used to investigate the frequency of incidents and to categorize similar incidents together.

The purpose of the interviews was to address the teacher and students’ recognition of norms, and their position (alignment) with the norms. The teacher interviews consisted of one semi-structured interview, conducted during the second week of the school year, and multiple informal interviews. The interviews included questions such as which norms are important and why, how the teacher wanted to initiate and operationalize the norms, and which norms the teacher thought were sustained and appropriated. The students' interviews consisted of a semi-structured small group interview, conducted during the sixth week of the school year, and informal individual interviews. The interview supposed a situation in which the group was asked to help a new student, named Jenny. Questions included "Jenny wants to know some of the rules of your science class so that she can actively contribute to discussion. What are three rules? About the norm that you guys just mentioned, do you think that you are doing well or not?," and "Why do you think that the norm you've just introduced to Jenny is important to make discussion better in your science classroom?" The interview data were fully transcribed.

## Findings

### Classroom norms: participation and argumentation norms

Classroom norms were identified inductively first, and then categorized by referring to Cobb et al.'s work (1996, 2001) on classroom norms and participant structure. For example, a candidate norm was labeled a norm when the candidate norm appeared more than three times. In addition, I looked for instances indicating violation of norms as evidence of established norms for the classroom (Cobb et al., 2001). The classroom norms are basically divided into two categories: participation norms and argumentation norms (Table 1).

Table 1: Emerged participation and argumentation norms

| Category | Norm | Description | Example (T: teacher, S: student) |
|---|---|---|---|
| Participation norms (PA) | Listen to each other | Ask students to listen to what other people are saying | T: We need to be able to listen. It is very important. When you work in groups, you also need to listen very carefully so that you can really understand what your friends are trying to say.<br>S: Listen. We are supposed to listen in science class.<br>S: Tell me what I just said.<br>S: You've got to listen. |
| | Talk aloud | Ask students to state their opinions aloud | T: Not everybody may hear important questions and answers if you talk like this. Imagine when you are yelling at your dog. How would you call the dog? Really project your voice so that others can hear. |
| | Attention behaviors | There are short phrases used to turn the class's attention to the teacher. | T: Raise your hands to hear my voice. One, two, three, eyes on me. |
| | Hold for a minute | Often the teacher blocked a few, more available students' answers right after the questions or comments in order to give others opportunities to talk. | T: Now let's just hold for a minute so that others have a chance to think. |
| | Call on others to help you | When someone wants to continue, or dominate the discussion, the teacher asks her/him to call for others (from the class) to add comments. | T: Why don't you ask for help from your group? Are there any comments from Group 2 about Maximo's comments? |
| | Talk to your group/ class, not me (Talk to us, not just one member) | Some students tend to talk to the teacher, not to the class. Then the teacher asked the students to talk and explain their opinions to the class. | T: You talk to me; that is not good. You should talk to the class. |
| Argumentati on norms (AR) | Think, talk, and act like scientists | The teacher highlights that 'we are scientists.' She frequently used this phrase, especially when students conducted experiments. | T: As scientists do, we fill out sheets to record our data. Scientists like measuring.<br>S: We did that three times because if we just did once, we wouldn't know whether it was accurate. So we did it three times like scientists. |
| | No right or wrong answer | The teacher also highlighted that there is no right answer in science. | T: I heard something interesting because we did not hear that before. I'd like to you to continue. There is no right or wrong answer. We want to hear what you think and believe. |
| | Feel free to have other opinions (what do you think) | Encourage students to disagree with others and provide alternative ideas | T: Isn't this an issue if there is just one person for one opinion? Do we have to be embarrassed if we make a different choice from the others? No, we don't have to. |
| | Back up your claim with evidence (why do you think what you think) | Encourage students to explicitly explain how evidence is related to claims | T: Don't forget. You also need to think about how you convince the rest of your group, so you need to give them some evidence about why your idea is a good idea. |
| | Convince others of your opinions | The teacher explained several ways of convincing others and then encouraged students to convince others | T: We'd like to let others know what we know. How do we convince others? |

Because the teacher has a proactive role that introduces and initiates norms in the classroom, I interviewed Ms. Green, and asked her which norms she thought were important for fostering argumentation and how she planned to initiate and operationalize the norms. She reported her aim was to emphasize norms to evoke broad participation, especially for the first two months. The reason for this was her acknowledgment of the existence of unequal participation and power dynamics among students. As a mixed-grade class, five or six of the fourth-grade "oldtimers" dominated classroom talk and work in their groups. She expected that the stressing of participation norms might contribute to encouraging other members, especially those who were new to the class, to talk and participate more. For example, when she was asked to introduce the most important classroom norm, she chose “listen to each other.” According to her, the norm not only focuses on the importance of listening to others' opinions but also helps students to ensure understanding of others' opinions and equal distribution of opportunities among students by encouraging them to ask for others' opinions. She felt this enabled the broadest student participation.

In addition to directly asking students to listen to what others are saying, she also encouraged this listening norm by having students rephrase, interpret, or summarize what another student had just said. Interestingly, she also emphasized that some participation norms could provide a basis for fostering argumentation. She said that if students are not willing to listen to others' opinions, then there is no reason to argue something because they are just satisfied with their own opinions.

Regarding the argumentation norms, the norms she answered seem to be derived from her beliefs about science. She projected an image of scientific knowledge as changeable, depending on what evidence is used. She stressed that scientists explore and work hard to better understand the world. There is no such thing as a right or wrong answer; it depends on how you convince others, by using evidence. Hence, the frequently used phrase, *we think, talk, and act like scientists*, implies argumentation norms, because she thinks that the central part of scientists’ work is to make a theory (claim) and back it up with evidence to convince others. Because there is no right or wrong answer, feeling free to have opinions and making an effort to convince others with your evidence was also stressed.

## Negotiation of norms

Students come to understand and share a norm through ongoing negotiation, not simply by having a norm presented to them (Yackel et al., 2000). In Ms. Green's classroom, this negotiation process included clarifying the classroom's understanding of the norm, generating new related norms, and extending the meaning of the norm. In the following example, Ms. Green and students began their discussion about a norm, *listen to each other*, and continued to discuss why *what do you think?* is a good question. Then the discussion was extended, and they agreed to include “*why do you think*” questions as a new norm. Ms. Green said, “Remember how important it is that the groups listen to each other. I heard somebody say, ‘What do you think?’ I think that this is a really good question to encourage listening to each other.” She asked the students why they thought this is an important question. Marisa answered, “In this way, we can gather what other people think. Maximo said, “It can get others involved in the discussion.” He also pointed out that this question was good for opening up a discussion, so that group members can focus. Oscar added the point that scientists are really curious, so this question encourages a scientific mindset. Then Josue pointed out that this question was nicer than other questions:

| | |
|---|---|
| Josue: | Because it is nicer, instead of screaming like “c'mon” because you don’t really have a chance to talk because it is difficult to say you should participate. |
| Ms.Green: | It’s better than an invitation. So, you think *what do you think* is better than *why don’t you work*? Or *why don’t you do something*? Or *you should participate*. Why do you guys think it is better? |
| Maximo: | I agree with Josue. Because it also very hard to make them into something by saying to them. You can also say *why do you think*, and then they would also go along with the evidence. |
| Ms.Green | Good. Excellent comments. That goes along with evidence and backing up claims. So we will ask, what do you think and why do you think? |

As seen from this example, the importance of listening norms was related to asking the question, *what do you think?* The *what do you think?* question was related to the ‘think, talk, and act,’ like scientists norm according to Oscar’s comments, and it seemed to enable the class to address an important argumentation norm (‘go along with evidence’) by including the ‘why do you think’ question, as addressed by Maximo and Ms.Green.

After this discussion, when Ms.Green visited a small group, she often checked whether the group members conformed to or violated this norm. She asked, “So, did you ask Camila a question, a ‘what do you think and why do you think’ question?” Then Josue asked a question about whether all members should ask this question of each other. She answered, “Yes, I’m strongly encouraging you guys to do that. Scientists like asking about their opinions, and they really like asking question ‘why do you think that?’”

## Change of participation and power dynamics, argumentation within groups

In the whole-classroom-level interaction, the distribution of the students' participation seemed to expand over the six weeks of observation. Ms. Green said in the fourth week, "It is true that there are some boys and girls who used to dominate the class, but now some other people are ready to join. Teena will pitch in. Marisa is another one. Sofia is a really good thinker." As shown below, students' participation seemed to be more evenly distributed by the sixth week, and there were some students who participated significantly more in the whole classroom discussions (Figure 1).

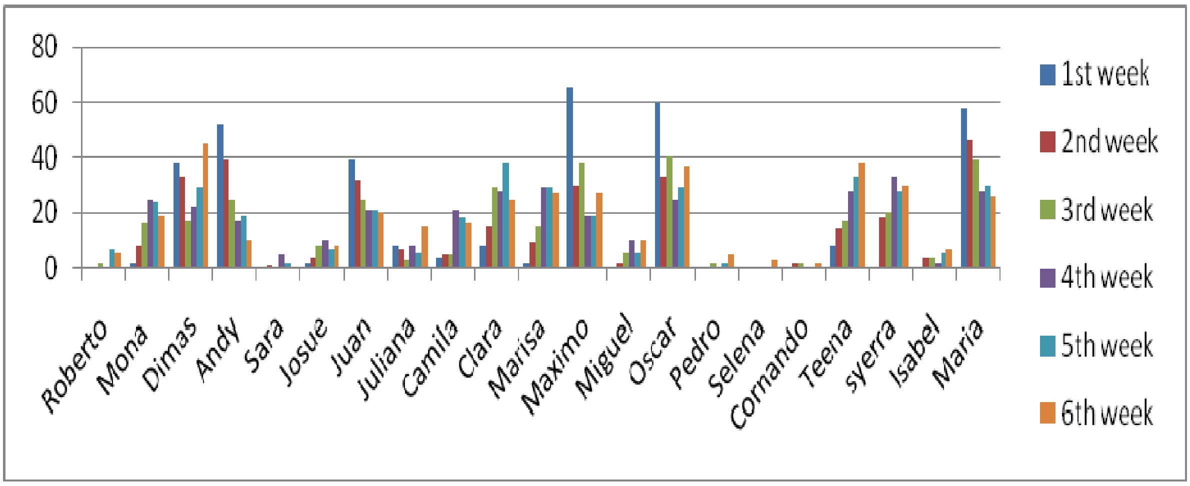



Figure 1. Distribution of individual students' participation through the first 6 weeks of class.

It must be noted that Ms. Green has distributed opportunities to participate on purpose, and often blocked the dominating students and had others talk. In other words, the improved distribution could be a result of the teacher's explicit manipulation of participation. Note that some of the frequently addressed norms, such as 'hold for a minute,' 'call others to help you,' and 'talk to your group/the class, not the teacher,' discourage students from dominating class discussions. For example, when Dimas said in the first week, "Ms. Green, I think…" in a little voice, the teacher asked him to turn around to the class and talk to the class, not the teacher only. She said, "Dimas, I know what you mean, but you should talk to your friends, not me. Turn around, stand up, and really project your voice to the class." Therefore, in order to see whether the establishment of norms enabled students to participate more, and encourage them to reorganize their norms, it was necessary to trace back to the changes in participation to small groups. The groups, organized by the teacher, are heterogeneous in terms of sex, grade, and academic achievement.

In order to see the changes in participation at the small-group level, I selected two groups for analysis. Each group included a member who showed significant improvements in participation at the whole-classroom level to see if the members who improved in whole class discussion also showed improvement in small groups. The analytical points of the group discussions included the participation frequency of each member, identifying the norms addressed during the discussion, recognition of students' roles (Herrenkohl & Guerra, 1998) and power dynamic, and counting and characterizing argumentation (Ryu & Sandoval, 2008) (Table 2).

Table 2: Analysis of small group discussion

| | Group 1 (Oscar, Clara, Pedro, Camila) | | Group 2 (Andy, Marisa, Teena, Juan) | |
|---|---|---|---|---|
| | First week | Sixth week | First week | Sixth week |
| Taken assigned roles (S: starter, R: recorder, H: helper, G: gatherer) | Oscar: S, Clara: R<br>Pedro: H, Camila: G | Oscar: R, Clara: S<br>Pedro: G, Camila: H | Andy: S, Marisa: R<br>Teena: H, Juan: G | Andy: R, Marisa: S<br>Teena: G, Juan: H |
| Taken audience/ intellectual roles (C: challenger, CL: clarifier, F: facilitator H: hypothesizer N: negotiator, S: summarizer) | Oscar: CL, H, S<br>Clara: H<br>Pedro: none<br>Camila: none | Oscar: CL, H, N, S<br>Clara: C, F, H, N<br>Pedro: CL, S<br>Camila: F, S | Andy: C, CL, H, S<br>Marisa: F, S<br>Teena: C, CL<br>Juan: C, H | Andy: C, CL, F, H, S<br>Marisa: C, CL, F, H, S<br>Teena: CL, N, S<br>Juan: H, S |
| Addressed participation norms (frequency) | Talk aloud (2) | Listen to each other (2)<br>Talk aloud (1)<br>Call others to help you (1) | Talk aloud (1) | Listen to each other (4)<br>Talk aloud (2)<br>Call others to help you (1)<br>Hold for a minute (1)<br>Talk to us, not just one member (1) |

| | | | | |
|---|---|---|---|---|
| Addressed argumentation norms (frequency) | Like scientists (1) | Like scientists (1)<br>Feel free to have other opinions (1)<br>Back up your claim with evidence (2) | Like scientists (1) | Like scientists (2)<br>Back up your claim with evidence (1)<br>Convince others of your opinions (1) |
| # of arguments | 0 | 2 | 0 | 3 |
| # of arguments with warrants | 0 | 1 | 0 | 1 |
| # of arguments with rebuttals | 0 | 2 | 0 | 2 |
| Power dynamic | Oscar is a dominant member, a main decision maker, and a task performer. | Clara increased her power as a challenger and facilitator. | Andy and Juan are main decision makers and performers. | Marisa and Teena's power increased a lot. With equal power, this group seems to have more arguments. |

Let us look at the following example. Ms. Green asked the following question, and had the group members discuss it: what will happen to two magnets if the space between the two magnets is increased?

The second week 10/06/09
Oscar : (He is counting the numbers of washers on the table, and is looking at the group's worksheet.)
Clara: Camila, do you know what will happen? (She asked the question of Camila, and looked at Oscar.)
Camila: (She does not say anything back to Clara, and turns to Pedro.)
Pedro: Think, think, think!
Ms.Green: Has any group come up with an idea about what will happen? Oscar?
Oscar: It gets weaker.
Teacher: What gets weaker? Clara, can you add some comments to Oscar's?
Clara: (She looks at Oscar. But Oscar does not say anything and looks atMs.Green.)
Ms.Green: Oscar, did your group discuss this?
Oscar (He does not answer the question, and he continues what he just mentioned.) If I put one, it is still strong. But if I put two, it gets weaker. The magnetic force is then weaker.

As seen above, Oscar, a fourth-grade male student, seemed to know or had an idea about the teacher's question, but he did not discuss it with the others. Even when Ms. Green asked Clara to add some comments, and she seemed to want to ask for help from Oscar (although she did not explicitly ask for help), he did not respond to Clara, and answered the question by himself.

In this group (Group 1), there was no significant discussion among the group members for the first three weeks. When there was a question or mission given to the group, Oscar made most of the decisions by himself, and conducted most experiments, even though there were roles assigned by the teacher, such as starter, gather, helper, and recorder. While he was not very talkative, he seemed to hold a dominant position in this group because he took on the roles of the sole decision maker and the main task performer. Oscar usually provided a claim but often without backing it up. Of the other members, Clara, a fourth-grade student who made improvements in participation, did not contribute to the discussion that occurred during this time. Pedro, and Camila, both third-graders, did not make additional contributions after Oscar's claim.

The other group (Group 2) consisted of Juan, a fourth-grade male student, Teena, a fourth-grade female student new to the class, Andy, a third-grade male student, and Marisa, a third-grade female student also new to the class. The girls are newcomers in this school. Juan and Andy talked to each other, answered questions together, and did experiments. While Marisa often provided her ideas, especially when she was assigned as a starter (who is supposed to explain the goal for the experiment), the two boys did not really respond to her ideas.

The second week 10/07/09
Marisa: How thick is different. Magnetic force is different because how thick the desk is and twenty pieces of paper are different.
Andy: Doesn't really matter. Go get the compass.
Marisa: Doesn't it matter how thick it is?
Andy: It doesn't matter.
Marisa: Fine. It doesn't matter.

Now look at the following two examples of the groups that occurred in the sixth week. The following discussion happened in Group 1 after the students generated their graphs based on their T charts: Oscar and Clara talked about whether they wanted to finish the graphs so that they could present them or just show the class the T chart showing the number of washers used to separate two magnets.

The sixth week 10/28/09
Oscar: So, you said that graphs are better. *Why do you think* that a graph is better?
Clara: Graphs told us information.
Oscar: But it is the same info.
Clara: No. It's not. This one (T chart) tells like numbers, how much it is, and the graph shows, the graph shows

| | |
|---|---|
| | shapes, and lines like. |
| Camila: | So we just want to show two lines? |
| Ms.Green: | (Ms.Green was listening to Oscar and Clara talking.) Can you see a pattern from here? (Pointing to the T chart) |
| Oscar and Clara: | No. |
| Ms.Green: | Can you see a pattern from here? |
| Oscar and Clara: | Yes. |
| Oscar: | I see. We cannot remember information from the T chart, but we can see and remember from the graphs. I like your idea. Let's just finish the graphs. I'll get stickers so we can finish. |

In this dialogue above, Oscar conformed to the norm, asking a 'why do you think' question. When Clara claimed that graphs told us information, Oscar disagreed and provided a rebuttal that it is the same information. Clara disagreed with his claim again, and provided a further explanation of her claim, why she thought that T charts and graphs are different. Although Clara did not fully support her ideas, both Oscar and Clara understood the difference between the two when the teacher provided a hint, and Oscar accepted Clara's opinion at this time. In addition, a noticeable difference found is Oscar's use of 'we' as the subject. In the first weeks' example, Oscar tended to use "I" whenever he talked. The use of 'we' seems to reflect Oscar's newly established group membership.

Let us look at the example of Group 2. This group conducted an experiment regarding how many washers are necessary to separate two magnets. The group's members talked about whether they needed to apply one washer at a time or more than one. Marisa argued that putting one at a time was a more accurate method, and Andy argued against her, because he thought that it could waste time. While he argued that they might need to put one at a time at the last measurements, Marisa disagreed with him because it could be difficult to figure out which would be the last measurement if they applied multiple washers.

| | |
|---|---|
| The sixth week | 10/26/09 |
| Marisa: | If somebody put five at a time, maybe the balance is five, but maybe the balance is two and if you put five, then maybe you need to pick some out, more in to see find balance, then it takes much longer than putting in one at a time. |
| Andy: | I disagree. Because when you put it in there, it doesn't matter how many you put in as long as you are going to put. Whether you put two or one, it doesn't matter until last time you put. So you put five at a time. |
| Marisa: | I don't understand. See we are doing like, by two, and then broke. And it can be seventeen or eighteen. It was seventeen? Or eighteen? Then you didn't quite know which one was. |
| Andy: | Look. If it breaks, then just go back to the last one. Then you need to put one at a time. Just do last time, and add one at a time. |
| Marisa: | One more what? |
| Andy: | Because it has to be one for the last time. But not for other times. It wastes time. |
| Marisa: | What if we broke the seventeen, but you put two? So it can't to make it, actually it wasn't seventeen, you don't know. |
| Andy: | But you know that it's not sixteen if you put two at a time. It is seventeen or eighteen. Right? |
| Marisa: | Yeah, I understand what you mean. But it is harder than putting just one at a time, then you may get the block. You know it just not accurate. |

A clear difference between the first week discussion and this week's is Andy's way of talking. While he did not explain why thickness did not matter in the example from the first week, now he explained in much more detail. Marisa's behavior also differs from that of previous weeks. In the first week, she said, "Fine." In the sixth week, she said, "I don't understand," and elaborated her claim. Although these two students did not reach a consensus, the way that they talked to each other has improved.

## Concluding Remarks

From the findings above, while the establishing of norms seems to work to some extent to encourage students' participation and argumentation, it is difficult to make a generalization that norm establishment can make every single student in the classroom participate more, and argue more. There are still silent students. However, for those who made improvements in whole classroom discussion such as Clara and Marisa (showed great improvement in their small group discussion), their participation seemed to enable the group to have more discussions. This indicates that the establishment of norms, guided by the teacher's discursive practice, improved the distribution of participation.

The teacher, Ms. Green, highlighted the norm "listen to each other" as a way to come up with new related norms, to interpret or extend the meaning of the norm through the negotiation process. Using this norm, students became more likely to monitor each other's thinking and ask about their reasoning to back up their claims. In other words, the students actively took audience roles. Taking these audience roles, in turn, seemed to help the students to take more intellectual roles because asking questions such as what do you think or why do think often became the same returning questions, what do YOU think and why do YOU think.

Hence, the findings suggest that paying more attention to participatory norms and roles may help students to appropriate argumentation norms and intellectual norms, and thus encourage students to have more scientific argumentation. While we often think that different opinions or disagreements are generated only from cognitive conflicts, which lead to scientific discussion, they could be triggered by social ones linked to cognitive ones. From this perspective, dynamic power relationships caused by social and cognitive conflicts may be encouraged by asking students to take on audience and intellectual roles.

## References

Andriessen, J. (2006). Arguing to learn. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 443-459). Cambridge: Cambridge University Press.

Brown, A. L. (1992). Design experiments: Theoretical and Methodlogical Challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141-178.

Brown, A. L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J. C. (1993). Distributed expertise in the classroom. In G. Saloman (Ed.), *Distributed cognitions:Psychological and educational considertations* (pp. 188-228). Cambridge, MA: Cambridge University Press

Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking, or a contex by any other name *Human Development, 21*, 108-216.

Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in Classroom Mathematical Practices. *The Journal of the Learning Sciences, 10*(1/2), 113-163.

Cobb, P., & Whitenack, J. W. (1996). A method for conducting longitudinal analyses of classroom video recordings and transcripts. *Educational Studies in Mathematics, 30*(3), 213-228.

Cornelius, L. L. (2008). *Student Participation in Desciplinary Discourses:When the Teacher Takes a Step Back, Who Takes a Step Forward?* Paper presented at the International Conferences of the Learning Sciences, Utrecht, Netherlands.

Creswell, J. W. (2003). *Research design Qualitative, quantitative and mixed methods approaches.* Thousand Oaks,CA: Sage.

Driver, R., Asoko, H., Leach, J., Scott, P., & Mortimer, E. (1994). Constructing Scientific Knowledge in the Classroom. *Educational Researcher, 23*(7), 5.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287-312.

Duschl, R. A. (2008). Science education in 3 part harmony: Balancing conceptual, epistemic and social goals. *Review of Research in Education, 32*, 268-291

Erduran, S., & Jiménez-Aleixandre, M. P. (Eds.). (2008). *Argumentation in science education: perspectives from classroom-based research.* Dordrecht: Springer.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education, 88*(6), 915-933.

Erickson, F. (1992). Ethnographic microanalysis of interaction. In M. D. LeCompte, W. L. Millroy & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 201-225): Academic Press

Herrenkohl, L. R., & Guerra, M. R. (1998). Participant Structures, Scientific Discourse, and Student Engagement in Fourth Grade. *Cognition and Instruction, 16*(4), 431-473.

Merriam, S. B. (1998). *Qualitative research and case study applications in education* San Francisco, CA: Jossey-Bass

Miller, D. C., & Salkind, N. J. (2002). *Handbook of Research Design and Social Measurement* (6 ed.). Thousand Oaks, CA.: Sage Publications.

NRC. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: National Academy Press.

Rosebery, A. S., Warren, B., & Conant, F. R. (1992). Appropriating Scientific Discourse: Findings From Language Minority Classrooms. *The Journal of the Learning Sciences, 2*(1), 61-94.

Ryu, S., & Sandoval, W. A. (2008). *Interpersonal influences on collaborative argument during scientific inquiry* Paper presented at the American Educational Research Association

van Eemeren, F. H., Grootendorst, R., Henkemans, F. S., Blair, J. A., Johnson, R. H., Krabbe, E. C. W., et al. (1996). *Fundamentals of argumentation theory:A handbook of historical backgrounds and contemporary developments.* Mahwah, NJ: Lawrence Erlbaum

van Zee, E., & Minstrell, J. (1997). Using Questioning to Guide Student Thinking. *The Journal of the Learning Sciences, 6*(2), 227-269.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes.* Cambridge, MA: Cambridge University Press

Webb, N. M., Nemer, K. M., & Ing, M. (2006). Small-group reflections: Parallels between teacher discourse and student behaviour in peer-directed groups. *The Journal of the Learning Sciences, 15*(1), 63-120

Yackel, E., Rasmussen, C., & King, K. (2000). Social and sociomathematical norms in an advanced undergraduate mathematics course. *Journal of Mathematical Behavior, 19*(3), 275-287.

# Interactional Arrangements for Learning about Science in Early Childhood: A Case Study Across Preschool and Home Contexts

Siri Mehus, University of Washington, smehus@u.washington.edu
Reed Stevens, Northwestern University, reed.stevens@northwestern.edu
Linda Grigholm, University of Washington, lgrigholm@gmail.com

**Abstract:** In this case study, we compare activities experienced by one 5-year old child in her home and her preschool in terms of how they afford scientific exploration and inquiry. We focus specifically on two types of interactional configuration: the first is an adult-guided mode of interaction in which a parent closely monitors the child's attention and action and accordingly orchestrates the emergence of opportunities for scientific observation, exploration and knowledge construction; the other is a type of peer play that occurs frequently at preschool, in which children collaboratively explore the physical properties of the world around them. Through close microanalysis we compare these activities as *learning arrangements* (Stevens, Satwicz & McCarthy, 2008), considering how they allow for or constrain children's entry into scientific activities, as well as the particular opportunities for learning they may afford.

## Introduction

In this paper, we compare aspects of the home and preschool contexts in terms of how they afford particular types of scientific exploration and inquiry in early childhood. Rather than surveying the entire range of activities in which children may engage in each of these environments, we focus on two types of interactional configuration which map onto these contexts—by this we mean not just that we have found them occurring in one of these settings, but rather that there are particular aspects of the environments that make them more likely to occur in one than the other. The first is an adult-guided mode of interaction in which the adult closely monitors the child's attention and action and accordingly orchestrates the emergence of opportunities for scientific observation, exploration and knowledge construction. Like others (e.g., Goodwin, 2007; Crowley & Jacobs, 2002), we have observed this occurring in family interactions. The other configuration is a type of peer interaction that occurs frequently within "free-play" periods at preschool. Often in the service of other goals, which may be fluid and loosely defined, children collaboratively explore the physical properties of the world around them—making observations, developing hypotheses, and even conducting informal experiments in short-lived social groupings of varying sizes. In both contexts, the type of science learning we focus on in this paper might be termed emergent or serendipitous (e.g., Stevens, 2000; Bell, et al., 2009) in that it occurs in the process of everyday living and playing with others, rather than within events that are explicitly designed for science learning.

In the pages that follow we analyze several interactions engaged in by one 5-year-old girl in our study. These include home interactions in which the child and her parents discuss the plants growing in their garden and the child's techniques for watering them, as well as preschool-based peer interactions such as one in which the child and her peers use a tape-measure to calculate the length of various objects in their classroom and then discover how to use it as a projectile. Through close microanalysis we compare these activities as *learning arrangements* (Stevens, Satwicz & McCarthy, 2008), considering how they allow for or constrain children's entry into scientific activities, as well as the particular opportunities for learning they may afford.

## Background

This study responds to previous calls for research investigating learning across school and out-of-school experiences (Stevens et al, 2005). While much educational research has focused on children's school experience at the expense of what happens in the home, research on informal science learning in early childhood has focused primarily on the family. Such research has shown that children have opportunities to learn scientific concepts in everyday life, and that adults in their lives have practices for supporting them in this process (e.g., Tizard & Hughes, 1984; Callanan & Oakes, 1992). Other researchers have pointed to the value of learning from peers in early childhood (Rogoff, 1990; Williams, 2001). Our observations suggest that informal play activities create a rich interactional environment for such peer learning. Interestingly, a primary site for such informal peer interactions seems to be the quasi-formal context of the preschool.

While educational research on school-age children may tend to focus on the classroom more than the home, research on learning and development in early childhood has tended to focus on the home, and particularly the dyadic interactions between parents and children that occur there. This tendency may be partly due to dominant cultural notions about child-rearing in the American middle class, but has been compounded through the research methods usually employed to study interaction and development in early childhood (Ochs

& Schieffelin, 1984). Our ethnographic approach, on the other hand, allows us to examine both home and preschool settings in their own right to identify and examine the ways in which learning is occurring, which can then be compared with the other contexts of children's lives.

## Study Description

This paper reports on one piece of a larger study examining young children's learning (2-5 years old) as it occurs in the multiple environments of their everyday lives, environments which may differ markedly in the opportunities for participation and resources for learning they offer. By observing and recording young children's interactions in multiple contexts (such as preschool, home, and playgroups) and among different configurations of co-interactants (including peers, older and younger children, teachers and caregivers), we seek to capture the complexity and variety of the social environments in and through which young children learn.

The first phase of this study involved weekly video-recording in a childcare classroom for approximately five months. In addition, two of the children in the classroom were also observed and recorded in their home environments. All in all, approximately 98 hours of video were collected, digitized and logged. In a second phase of this study, we are video-recording in four rooms at two preschools for a total of nine months. In addition, out-of-school activities of eight children are being recorded.

In this paper we report on the first phase of the study, in which we observed and recorded activities in "Rocket Room," a kindergarten-readiness classroom (4-5 years old) in a nonprofit NAEYC-accredited childcare center. The center was located in a middle class neighborhood of a mid-sized Western city. There were 20 children in the Rocket Room, supervised by a lead teacher and an assistant teacher. Our observations took place in the mornings, during which time the children participated in free-play, breakfast, art activities, and circle time. We describe learning interactions experienced by Darcy, one of the focal children in the study, both in the preschool and her home environment. Darcy was 5 years old at the time of the study and had an 18-month old sister. Both Darcy and her sister attended the childcare center full time. Darcy's parents both worked in high-tech occupations.

We employ the method of video-based microanalysis of interaction (Stevens & Hall, 1998; see also Jordan & Henderson, 1995; Streeck & Mehus, 2005). Our moment-by-moment analysis allows us to demonstrate possibilities for learning and requirements for interaction in the activities in which children participate. This form of qualitative inquiry allows us to cautiously generalize our findings not to populations but rather to practices and activities based on principles derived from analyzing structures of social action.

## Analysis

### Parent-Guided Learning Interactions at Home

When video-recording children in home environments, we frequently captured interactions that were intensely mediated by parents. In the section that follows, we describe some such interactions that arguably could have relevance for the development of scientific skills and interest, and analyze the particular opportunities for learning they afford.

#### Key Characteristics of the Learning Arrangement

What do we mean by "parent-guided learning interactions"? These are events that may be planned in advance, but even if not they are organized for children's learning as they proceed. Within these activities, children's talk and actions are closely monitored by adults and frequently commented upon, usually praised. Children's actions and other aspects of the environment are frequently used as a starting point for elaboration and exploration, within which there are clear (and often explicit) attempts to guide children's perception toward particular aspects of the scene and to provide labels for what children are experiencing (Stevens & Hall, 1998). The segment below exemplifies several of these features. This occurs in the back yard of Darcy's home. Darcy is splashing in an inflatable swimming pool while her mother and father watch. Her mother gets up to take a look at their garden, and calls Darcy's attention to what she sees.

Table 1: Data Excerpt

| | | |
|---|---|---|
| 1 | Mom: | Here you know what (.) Darce? It looks like the (.) the: uh sweet peas (over here) |
| 2 | | need water. (That soil still) looks dry: doesn't it. |
| 3 | | *((Darcy starts watering where her mom pointed))* |
| 4 | Mom: | Do you need some help with that? *((gets up))* |
| 5 | | It's probably on the far side so it's kinda (.) kinda rough huh. |
| 6 | Mom: | You want me [to |
| 7 | Darcy: | [I can smell the sweet peas. |
| 8 | Mom: | You can smell the sweet peas? What do they smell like. |
| 9 | | Do they smell (.) sweet? |
| 10 | Darcy: | *((grunts))* |
| 11 | Mom: | Woah. (.) |
| | | *((Darcy is rocking watering can back and forth))* |

| | | |
|---|---|---|
| 12 | | That's kinda clever. |
| | | *((Darcy steps up onto side of bed))* |
| 13 | Mom: | You're like (.) rocket-shipping. |
| | | *((Darcy reaches out with watering can over bed, wobbles))* |
| 14 | | Oh there you go. |
| 15 | | Here I'll hang on to you: *((Mom holds Darcy around torso. Darcy holds watering can))* |
| 16 | Mom: | Hang on to you: (.) |
| 17 | | That's kinda heavy huh. *((takes watering can))* |
| 18 | | °Here you go. ° |
| 19 | | (can I) do these? |
| 20 | Darcy: | Mm hm. *((turns away from garden, towards pool))* |
| 21 | | *((Darcy gets into pool, Mom continues watering))* |
| | | … [01:10 exchange between parents and researcher not transcribed] |
| 22 | Mom: | Darcy! (.) Did you no:tice? |
| 23 | Dad: | Oh there's some sprouts coming up? |
| 24 | Mom: | Check it ou::t! |
| 25 | | Holy cow that's fa:st. We just put this in last week. |
| 26 | | Look! |
| 27 | | What are tho::se? |
| 28 | Darcy: | Sprouts. |
| 29 | Mom: | What kinda sprouts. |
| 30 | Darcy: | Sunflower sprouts |
| 31 | Mom: | =sunflower sprouts. [Holy cow::! |
| 32 | Dad: | [°Sunflower sprou::ts.° |
| 33 | Darcy: | I've done (.) some (.) very good watering |
| 34 | Mom: | =You've done some very good watering. |
| 35 | | That is instant gratification. (.) [Right there. |
| 36 | Dad: | [Oh. Educational moment. |
| 37 | Researcher: | Yeah. |

This segment exemplifies several of the features that characterize adult-child learning interactions in the homes that we studied. First, these *events are managed for children's learning*. In some cases, they are pre-planned, involving the preparation of materials and the use of predesignated roles and actions. Even when they are not, the activities are organized by adults in the moment, with slots for children's participation being provided by the parents. While the particular event above was not planned at the level of a classroom lesson, it was purposefully brought into being for educational purposes by the parents. Darcy's mother told us that they decided to plant a garden (their first) specifically for Darcy's benefit (as a means to "inspire little Darcy about growing things," as well as potentially get her more interested in eating vegetables). Within the moment, Darcy's mother enlists her participation in a familiar interactional routine. For examples, in lines 27-31, Darcy's mother engages her in a school-like initiation-response sequence in which the mother is asking Darcy "known-answer" questions (cf. Mehan, 1979; Cazden, 1988).

Within these events, *children's actions and talk are closely monitored*, explicitly commented on, often praised, and elaborated upon. For instance, in line 12 Darcy's mother verbally calls attention to Darcy's action and labels it as "kinda clever," further describing it in line 13 as being "like rocket-shipping". *Physical phenomena in the environment are also noted and commented upon*, even used as opportunities for "occasioned knowledge exploration" (Goodwin, 2007). We see this in lines 1-3 above, when Darcy's mother first calls her attention to the garden's need for water, as well as in lines 22-34, when she discovers new growth and subsequently guides Darcy in specifying what they are seeing. Parents orient to these as potentially educational moments (though not usually as explicitly as does Darcy's father in line 36 above).

## Opportunities for Learning in Parent-Guided Interactions

There are several types of learning opportunities that children encounter in these types of interaction. Many of them are provided by parents through talk. For instance, a prominent source of potential learning in the interaction above comes from the mothers' directing of her daughter's attention and guiding or disciplining of her perception (e.g., Stevens & Hall, 1998). We see this in lines 1-2, in which Darcy's mother not only calls her attention to a specific aspect of the environment, but guides her in how to *see* it—i.e., to recognize the dirt as soil that "looks dry."

We can also see this in the way Darcy's mother calls attention to Darcy's actions. For example, in this exchange Darcy's mother verbally takes notice of an action being performed by the child and frames this action as being both purposeful and ingenious. In Line 12 an explicit positive assessment is made, linked to the child's action with the deictic pronoun "that's." This assessment is followed by elaboration of *why* the child's action seems like a good idea, in the form of a description of what the child is achieving through her action (line 13). At that moment Darcy is standing on the edge of the garden bed and rocking the watering can back and forth, which causes the water to slosh out, reaching further into the bed than it would if she were simply pouring it out. Darcy's mother's "rocket-ship" analogy conveys the notion that through her action, Darcy is causing something to be propelled through the air, "cleverly" solving the problem of how to water the far side of the garden. This represents a compelling example of the ways in which adults can frame aspects of the world for children, in this case quite possibly shaping a child's retrospective understanding of her own action.

Darcy's mother also engages Darcy's participation in these labeling activities. For instance, in line 27 Darcy's mother elicits a label from Darcy by asking "What are those?", and then requests specification in line 29 with "what kinda sprouts?". As Darcy participates in these activities, she potentially learns ways of seeing, categorizing and labeling the world around her—as well as ways of thinking about and valuing her own actions.

Finally, it should be noted that Darcy takes advantage of this interactional arrangement to initiate and further her own learning experiences. Twice she issues declarative statements that draw responses from her mother. For instance, in line 7, Darcy asserts that she can smell the sweet peas. Her mother responds with the question "what do they smell like?" then proposes a candidate answer: "Do they smell sweet?" In line 33, in response to her mother's enthusiastic pointing out of the new sprouts, Darcy comments that she (Darcy) has "done some very good watering." The timing of Darcy's assertion suggests that she is making a connection between the sprouting of the sunflower seeds and her previous watering actions—an idea that may have been planted (so to speak) in previous interactions with adults. By making this claim in the context of an interaction with her mother, Darcy makes it available for confirmation, which her mother provides unambiguously in the next line.

### Discussion: Adult-Child Interaction in the Home as Learning Arrangement

The participation framework of adult-child talk in the home makes for excellent learning opportunities for children: parents guide children's attention, deliver comments and questions that are tightly coupled with children's activities, and design and orchestrate interactions with children's learning in mind. Children's initiations of learning opportunities are taken up readily and enthusiastically. The interactional demands on children are low: adults do much work to invite children's participation and ensure that these activities work as learning arrangements. These do not represent the totality of children's interactions at home; there are times when the children do not have their parents' full attention. Even when they do, parents are not always able to take advantage of every learning opportunity. It should also be noted that other types of interaction that occur in this home (and other homes) may offer different types of learning opportunities (such as guided participation in adult activity, intent observation of adult or peer activity, peer learning in sibling play, to name only a few). We focus on this type both because we observed it frequently, and because particular features of the interaction format make it rich with learning opportunities, as described above.

We propose that structural and cultural aspects of the home setting make these types of interaction more likely to occur there than in preschool, most simply the numbers of adults and children. In preschool, the ratio of children to adults means that teachers must frequently shift attention in order to monitor and care for everyone. While adults in the home may also have other demands on their attention, they are much more likely to be able to carry out extended interactions in which they maintain focus on the talk and action of an individual child. This does not mean that similar interactions cannot occur in preschool settings, and in fact they do. It does suggest, however, that the unique features of the preschool environment may give rise to *other* interaction patterns, which can in turn be analyzed for the ways in which they allow science learning to occur.

## Learning through Peer Interaction in the Preschool Classroom

In this section, by way of comparison, we analyze the learning opportunities inherent in an interactional arrangement that we found in our preschool observations and did not find to occur in home—these are small group activities, organized and guided by children. Though we do not preclude the possibility that this type of activity might occur in other contexts, there are certain features of this environment that make it possible. For instance, during free-play periods at school, teachers rarely spend long periods of time with a single child or group of children. This means that while they do set up activities for children and step in to mediate children's interactions, teachers infrequently engage in the intense moment-to-moment guidance typical of the homes we observed. Rather than having interaction initiated and maintained for them, children engage one another.

### Key Characteristics of the Learning Arrangement

The type of activity analyzed in this section is characterized by (1) including 3 or more children (more children may come in and out of the activity), (2) being initiated, guided and maintained by the children, rather than by adults, (3) enduring over a relatively long stretch of time (for example, the measuring activity analyzed in this paper endured for nearly 45 minutes in total), (4) being organized towards some purpose (although individual goals may not be shared by all participants and may shift during the activity). For the purposes of this paper, we analyze one extended activity in which all four characteristics are present, as representative of this type of interaction format. Though the activity we describe is not explicitly undertaken as a "science" or "math" activity, there are opportunities for scientific and mathematical exploration and discovery available within it.

## Measuring Tape

The activity we analyze in this section occurs over an extended period of time during which a core group of three children, joined occasionally by other children in the room, play with a tape measure that one of the children has brought from home. It is free-play time in the preschool classroom. One of the children, Anna, gets up from a table and shows a drawing to one of the teachers (Nancy). She then brings it over to her cubbie, puts it in, and pulls out a tape measure. "I brought this from home:::!" she says enthusiastically to an entering parent. Anna then brings the tape measure over to Nancy who compliments her on the idea to bring a tape measure from home and asks her what she wants to measure. Anna begins to measure a bookshelf. Though Nancy soon becomes occupied with other children, Anna is able to regain her attention long enough to get some help with the measuring task she has begun. Specifically, Nancy provides instruction in reading numbers from the tape ("A three and a seven is thirty-seven.") and with the unit of measurement ("An inch is about that long," gesturing with thumb and forefinger). She also provides a different type of guidance: in the form of a question, she suggests to Anna ways of linking her current measuring activity with adult professional practices ("Are you a carpenter? A mathematician? A scientist?"). After this point, the children's measuring activity proceeds with no significant input from any of the adults in the room.

Perhaps taking her cue from Nancy's question, Anna then initiates a game with some other children, in which Anna plays the role of carpenter. That game is short-lived, and when it ends Anna and one of the children (Mira) begin stretching out the measuring tape and letting it go so it retracts into the case. They are then joined by one of the girls who will make up the core group: Susan. Susan suggests that they measure people, and though Anna initially resists this idea, when we find them again (after a one minute gap in which they are out of camera range) Susan is measuring Anna's height, while Mira watches. Susan struggles with the challenge of holding the tape at the floor with one hand and extending the to Anna's head with the other, and then being able to see the numbers at the top.

It is at this point that Darcy (one of our focal children and the third girl forming the core group) joins them. She approaches the two girls, takes the end of the measuring tape, stretches it to the top of Anna's head, and holds it there. This allows the girls to move to the next step of the measuring task, which is to read the number. Anna and Susan then measure Darcy, and the three then move on to measuring a large cardboard tree. The three girls continue to work together as a team, measuring objects all over the preschool classroom.

For most of this period, they measure pre-existing objects, such as tables, the refrigerator, chairs, bookshelves, the elevated stage, room dividers, and a large block. During this time, they negotiate a set of roles and a turn-taking system. The set of roles evolves and shifts over the course of the activity, however, some core "jobs" emerge, such as holding (and hooking) the tab end of the tape measure; extending the tape measure case; looking at, reading, and shouting out the measurements (numbers); and being the one to yell "Let go!"; and releasing the tape measure case (so that it springs back to the hooked tab end as the tape retracts into the case).

In the last minutes of the activity, the girls begin to put blocks together to construct a low wall to measure. Though it may not be appropriate to attribute an overall purpose to the activity – as purposes surely shift locally and vary between individuals – a concern expressed explicitly and manifested in the girls' actions is to add a block between each measuring without reaching the point where the tape measure will no longer extend to the end of the block line. When the block line gets all the way to a low stage on one side of the room, they discover that when they let the tape case go it will slide along the top of the block line and then jump up onto the stage. This effect is greeted with celebratory whoops and jumps and repeated several times.

There is then an interruption of the dominant activity while the girls join some of their classmates in a different activity with one of the teachers. When the girls leave the line of blocks, other children immediately approach it and begin to appropriate the blocks for their own purposes. The girls enlist the help of a teacher (Nancy) in re-establishing the exclusive use of the block line for measuring activities. One of the children, Nathaniel, sticks around and gains access to the group by assigning himself a role ("I'm the guy who watches") in the activity. It is at this point that our focal child, Darcy, exits the group to engage in another activity.

## Learning Opportunities

Over the course of the period the three girls play with the measuring tape, there are multiple opportunities for them to learn from (and teach) each other. We focus here on the experience of our focal child, Darcy, and find types of learning afforded by the activity can be analytically separated into two rough categories. The first is more directly related to the activity of measuring as ordinarily carried out and relates to a set of concepts that include number identification, dimension (height vs. length), and quantitative comparison (testing whether one thing is longer or shorter than another thing). The other category is not directly tied to the activity of measuring, but rather to the physical affordances and constraints of the tape measure as an object. This set includes learning about the limitations of the material – both in terms of its length (the tape can only be stretched a certain distance) and in terms of its strength (the tape will collapse if it is stretched too far without any external support). Particularly interesting for Darcy and the other girls is a particular affordance of the tape measure tool: that the tape automatically retracts into the case when released, and that this mechanism will also cause the case end to spring toward the tape end when that end is held (or hooked on something). While the learning in the first category can be related by a knowledgeable observer to mathematics (including geometry), the second can be related primarily to physics and engineering.

Over the course of the activity we do not necessarily see a clear progression towards mastery of particular concepts or skills. However, we do see the children engaging in negotiation and discussion about how to do things, trying out terminology with one another, making discoveries, issuing predictions about the behavior of physical objects (informal hypotheses), and watching to see if their predictions are borne out.

*Learning about measuring*: One of the basic sub-tasks involved in measuring is *identifying and reading the numeral*s printed on the tape. Anna receives direct instruction in this from the teacher (as described above). After Darcy joins the group, the girls continue to negotiate how do to this. In the excerpt below, the girls are measuring their sixth object together, and still working out how to read numbers from the tape. Also under discussion and available for learning are issues of *dimension*; i.e., that there are different terms to describe measurements made in different directions (lines 13 and 15), and that this is a distinction worth noting.

Table 2: Data Excerpt

| | | |
|---|---|---|
| 1 | Susan: | *((jumps off stage and bounces over to Anna))* How about the ta::ble. |
| 2 | Anna: | (oew:::) |
| 3 | | *((girls run off-screen))* |
| 4 | Anna: | I know. |
| 5 | | How about we measure this table. |
| 6 | Susan: | Okay. |
| 7 | Anna: | Okay. so. |
| 8 | | Um Susan [we're measuring to thi::s |
| 9 | (Darcy:) | [(we're measuring the table) |
| 10 | Susan: | Okay. |
| 11 | Anna: | You know::? |
| 12 | | Okay um |
| 13 | Susan: | You mean you're measuring how long it is. |
| 14 | Anna: | Yes. |
| 15 | Susan: | Not how tall it is. |
| 16 | Anna: | Yeah. |
| 17 | Susan: | Hey. When it (springs) back it might (.) pinch your fingers. |
| 18 | Darcy: | Yeah. I know that. |
| 19 | Anna: | =How many [is it? |
| 20 | (Susan:) | [(so if it falls (.) back it might) |
| 21 | Darcy: | It's seventy-five feet. |
| 22 | Susan: | Lemme see! Actually (.) it's (.) fifty-seven feet. |
| 23 | Anna: | OKAY:: [(         ) |
| 24 | Susan: | [Actually actually it's (.) seventy-five feet. |
| 25 | Anna: | Okay. STA:ND BACK EVERYONE: |
| 26 | | *((sound of tape measure case scooting over table and dropping to floor))* |
| 27 | Girl: | YES! |
| 28 | Anna: | How about the fridge? |
| 29 | Susan: | Yeah we'll mea:sure the fri::dge! |

*Learning about the constraints and affordances of the object*: Over the course of their play activity, the girls encounter and explore various properties of the tape measure itself. For instance, they (particularly Anna) display awareness that the tape is limited in length, and concern themselves with limiting the length of their block wall accordingly. Earlier on, they discover that the strength, or the ability of the tape measure to support its own weight when extended, is also limited – e.g., when measuring a tall cardboard tree, the measuring tape bends over on itself and Anna yells, "It's too high!"). However, it is a curious affordance of the tool that captures their interest the most. That is the self-propelling action of the tape measure – that the tape end automatically retracts into the case, but also the case end can spring back to the tabbed end if the tabbed end is held in place (see lines 17-18 and 25-27 above).

That this is a learning experience for Darcy is evidenced at a later moment. The girls are measuring a bookcase by placing the tape measure case on the floor and extending the tape towards the top of the bookcase. When Anna lets go of the case, which is resting on the floor, Darcy abruptly pulls her hand back and jumps away from the tabbed end. Evidently, she expects the case to spring up towards the tabbed end (though it does not because it is too heavy). Darcy's reaction represents learning in progress. Her earlier experiences with the tape measure have led her to form the expectation that the case will always spring towards the tabbed end when released. It is fairly certain that she would not have expected a metal case to rise into the air if she had not had this experience with the tape measure. That the tape measure does not do what she expects provides an opportunity for Darcy to revise her conceptualization of the properties of this object, potentially gaining a more complex (though unarticulated) understanding of the interaction between gravity and the pull of the spring retracting the tape measure into the case.

Much later, after the girls have created a wall of blocks for measuring, the girls discover that when the tape measure is released so that it runs along the block wall, it will act as a projectile and pop up onto the stage (beyond the spot at which the end was hooked). The first time this happens, the girls react with excitement and try it again. The second try is less successful and meets with less enthusiasm. The third try is presented below.

Table 3: Data Excerpt

| | | |
|---|---|---|
| 1 | Anna: | OKAY. Now:::: *((nasal))* I wanna do (.) as many ( ) *((picks up blue block and puts at end of wall, looks back to Darcy* |
| 2 | | *and Susan))* |
| 3 | | *((Darcy stretches measuring tape along the length of the wall))* |
| 4 | Anna: | *((nods)) ((puts hands on hips))* I think (it'll) stop soon. Because it does stop. |
| 5 | | *((Darcy gets to end of block wall--back end of tape measure is lined up with end of wall))* |
| 6 | Anna: | Okay. *((looks at end of tape measure, looks at Darcy))* |
| 7 | (Darcy:) | Okay? |
| 8 | Susan: | Okay. |
| 9 | | *((Both girls look at Susan, holding tab end))* |
| 10 | Girl: | (All right. Let's go. Looks goo::d.) |
| 11 | | *((Susan lets go of hooked end and moves away from block wall.))* |
| 12 | Anna: | Okay [( ) |
| 13 | | [*((Darcy lets go of tape measure))* |
| 14 | | *((tape measure pops up onto stage))* |
| 15 | Anna: | Waaa[::oh *((runs toward tape measure))* |
| 16 | Susan: | [Wu Hoo::= |
| 17 | Darcy: | =Wuhoh *((jumps, walks to end of wall))* |
| 18 | | *((Susan jumps, walks toward group at table.))* |
| 19 | Anna: | Okay 'm gonna I'm gonna |
| 20 | | *((Susan moves chair))* |
| 21 | Anna: | *((calling out))* Okay. Put ano- another block the::re. |

Four times in all the girls let go of tape measure and it pops onto the stage. The girls display excitement about this when it happens (in the form of whoops, jumps, and screams). They do not make verbal reference to it. This activity is in some ways like an experiment. For instance, as she lets go of the case on the last try, Anna says, "let's see:::," which suggests that she is orienting to this as trying something out. It is also experiment-like in that it consists of multiple repetitions of a procedure with slight modifications each time. These multiple repetitions, in which the girls take different roles, allow them to see the phenomenon from different perspectives and observe which aspects remain stable and which aspects change. However, there is no evidence that the modifications undertaken are designed to produce an effect, especially not of making the tape measure jump when released. To the extent that there is a particular purpose indicated, it is to add blocks of the right number and length such that the tape measure will still extend to the end of the block wall (lines 1, 4, 19-21). Otherwise, the majority of the talk surrounding this activity is about the turn-taking process itself (e.g., "It's my turn.") and the appropriate performance of roles (e.g., "No. You have to say 'let go'!").

## Discussion: Peer-regulated Group Activity as Learning Arrangement

We can recognize several advantages of this form of interaction as a learning arrangement for young children. First, there appears to be strong social motivation for participating in these activities. Even when their participation is not solicited nor reliably rewarded by other children—as it tends to be by parents in the home environment—children actively seek out and work hard to gain entrance into these activities. We might say that the activities exhibit something like a gravitational pull on children. Furthermore, these activities afford different types of learning opportunities. Children negotiate with each other in guiding and regulating the activity themselves—as such they have increased opportunity to both exercise agency and to develop skills at collaborating with others. Children can position themselves as both learners and teachers in such interaction. The lack of stable and explicit goal orientation gives children a great deal of freedom to explore the aspects of a phenomenon that most interest them (Rogoff, 1990). These peer-organized learning arrangements also seem to offer enhanced opportunities for embodied participation and experiential learning, as children move about the rooms with little constraint, drawing into their activity a variety of different types of artifacts and structures.

There also are disadvantages when compared with adult-mediated interaction. These have to do with the interactional challenges of the activity and the lack of adult support in developing an activity in the direction of disciplinary knowledge. In child-regulated peer group activity the interactional demands fall squarely on the shoulders of the children themselves. Our analysis of this activity shows that gaining entrance is itself a difficult maneuver—while one our focal children succeeds, the other attempts to participate but is unsuccessful (even reprimanded for her attempt by the core trio of girls). Similarly, it is the children who must maintain the activity and direct it in fruitful and enjoyable directions. Young children may not always be able to do this effectively, which may be why such activities rarely last as long as the one we analyze in this paper.

Children also do not experience the same level of interactional support for learning in peer activities as they do in parent-guided activities. Unlike parents, peers may not monitor for opportunities for teaching and learning. To the extent that children's activities are monitored, peers are more likely to see shortcomings and provide correction than recognize innovation or respond with elaboration. These qualities of peer play do offer opportunities to develop skills at argumentation and negotiation--part of the collaborative practice of science.

Another possibility is that rich opportunities for learning may occur but not get developed in the direction of disciplinary knowledge. For instance, in the block-measuring activities described above there is no movement toward investigation of why the tape measure flies. Over the course of the repeated tries girls encounter 'information' about the physical workings of the tool; however, they do not seem to engage in purposeful or principled experimentation in order to find out more. Without the involvement of knowledgeable adults or peers there is little opportunity to link this activity to disciplinary principles or practices.

## Conclusion

The distinction between formal and informal learning takes on a peculiar status when applied to the lives of very young children. Though formal education reaches into the mostly informal lives of young children in many ways, "school" does not neatly map onto "formal" and "home" does not neatly map onto "informal". In fact, due to the particular social configurations of preschools (versus middle class homes), they seem to offer *more* opportunity for the highly informal learning arrangement of uninterrupted, child-organized group play.

In light of the affordances and limitations of informal peer-group interaction for science learning discussed above, it remains to be determined exactly what role participation in this type of activity might play in a trajectory of 'becoming a scientist' or even in building the disciplinary knowledge needed to succeed in science learning later in school. Our analysis supports pedagogical practices that take advantage of the opportunities for peer learning afforded by the preschool classroom setting by allowing children the freedom to self-organize into groups and pursue activities of their choice. On the other hand, our analysis also suggests that strategic adult support, provided at key moments in children's peer play, could provide disciplinary connections and promote deeper learning of a kind that may not happen when children play together without intervention.

## References


Bell, P., Lewenstein, B., Shouse, A. and Feder, M. (Eds.) (2009) *Learning science in informal environments: People, places, and pursuits.* National Academies Press.

Bruner, J. S. (1983). *Child's talk: Learning to use language*. New York: Oxford University Press.

Callanan, M.A. and Oakes, L. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development*, 7, 213-253.

Cazden, C. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Crowley, K. & Jacobs, M. (2002). Islands of expertise and the development of family scientific literacy. In G. Leinhardt, K. Crowley, & K. Knutson (Eds.) *Learning conversations in museums*. Mahwah, NJ: Lawrence Erlbaum Associates.

Goodwin, M.H. (2007). Occasioned knowledge exploration in family interaction. *Discourse & Society*, 18(1), pp. 93-110.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences*, 4(1), 39-103.

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization: Three developmental stories. In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self and emotion* (pp. 276-320). Cambridge: Cambridge University Press.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard U. Press.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. NY: Oxford U. Press.

Stevens, R., Satwicz, T., & McCarthy, L. (2007). In game, in room, in world: Reconnecting video game play to the rest of kids' lives. In Salen, K. (Ed.), *Ecology of games: MacArthur Foundation series on digital media and learning*. Cambridge, MA: The MIT Press.

Stevens, R. (2000). Who counts what as math: Emergent and assigned mathematics problems in a project-based classroom. In J. Boaler (Ed.), *Multiple perspectives on mathematical teaching and learning* (pp. 105–144). New York: Elsevier.

Stevens, R. & Hall, R., (1998). Disciplined perception: Learning to see in technoscience. In Magdalene, L. & Blunk, M. L., (Eds.), *Talking mathematics in school: Studies of teaching and learning*. Cambridge, MA: Cambridge University Press.

Streeck, J., & Mehus, S. (2005). Microethnography: The study of practices. In K. L. Fitch & R. E. Sanders (Eds.), *Handbook of language and social interaction*. Mahwah, NJ: Lawrence Erlbaum.

Tizard, B. & Hughes, M. (1984). *Young children learning*. Cambridge, MA: Harvard University Press.

Williams, P. (2001). Preschool routines, peer learning and participation. *Scandinavian Journal of Educational Research, (45) 4,* pp. 317-339.


## Acknowledgements

This work was supported by a grant from the National Science Foundation (NSF #0354453). Sheldon Levias and Lionel Enns contributed to the data collection.

# Promoting Learning in Complex Systems: Effect of Question Prompts versus System Dynamics Model Progressions as a Cognitive-Regulation Scaffold in a Simulation-Based Inquiry-Learning Environment

Deniz Eseryel & Victor Law, University of Oklahoma, 820 Van Vleet Oval Rm 321, Norman, OK 73019-2041 USA
Email: eseryel@ou.edu, vlaw@ou.edu

**Abstract:** Designing effective technology-based learning environments is challenging. Designing effective technology-based learning environments to facilitate learning about complex knowledge domains is more challenging. To a large extent, the key to the puzzle lies in identifying which scaffolding strategies are more effective; and under which conditions. In a simulation-based inquiry-learning environment, this controlled study investigated the effect of two promising scaffolding strategies; *question prompts* and *system dynamics model progressions*, on ninth-grade biology students' cognitive regulation and complex problem-solving skills. For simpler complex problems, findings suggested that both scaffolding strategies were equally effective. However, as the problems increased in complexity, system dynamics model progressions were significantly more effective for facilitating both cognitive regulation and complex problem-solving skills.

## Introduction

How can we effectively facilitate learning in complex knowledge domains such as science, technology, engineering, and mathematics (STEM)? This question led us to the study presented in this paper. Recent research on problem solving in complex knowledge domains illuminates a number of learning challenges due to the nature and structure of the real-life problems in these domains. For instance, ecology is a difficult subject matter for many students. Learning ecology calls for learning about the complex ecology systems. In other words, students should understand the dynamic interdependencies among different organisms and their environment. As such, a complex system involves a large number of interdependent variables and they are all connected with one another dynamically (Dörner, 1996). Prior research suggests that humans have significant difficulties in understanding complex systems, due to the challenges involved in building a mental representation of complex systems, predicting dynamics outcomes, monitoring the cause and effects of complex systems, and monitoring their own strategies of information processing. (Dörner, 1996; Dörner & Wearing, 1995; Funke, 1991).

Mental model theory provides a framework to understand the learning processes of complex systems, in which learners take the input from their learning environments and simulate the events in their mind to accommodate revised mental models (Rumel, Smolensky, McClelland & Hinton, 1986; Seel, 2001). While learning about complex systems, learners are required to construct and reconstruct their mental models, which (1) guide learners' comprehension of the systems, (2) allow them to explain the system states, and (3) allow them to predict the system behaviors (Greeno, 1989; Seel, 2006; Young, 1983).

Therefore, it is argued that simulation-based inquiry learning environments would help students in the inquiry processes, because they allow learners to visualize and investigate complex systems (de Jong & van Joolingen, 1998; Linn, Davis & Bell, 2004). In a simulated-based inquiry-learning environment, students are expected to discover dynamic relationships underlying a complex system through an iterative process of hypothesis generation, data collection, and data analysis. Indeed, simulation-based inquiry learning is gaining visibility, especially in the area of science education, as a way to address the challenges of complex learning (The National Research Council, 2000). Education researchers advocate simulation-based learning environments to foster scientific inquiry skills (Wilensky & Reisman, 2006). However, studies showed that many students have difficulties with simulation-based inquiry learning (Hogan & Thomas, 2001). In addition, the effectiveness of inquiry-based learning is debatable (de Jong & van Joolingen, 1998; Linn et al., 2004). Recent research suggests that low-level cognitive regulation skills are a major cause of failure in simulation-based inquiry learning (Lavioe & Good, 1998; Simmons & Lunetta, 1993; Schauble et al., 1991; Shute & Glaser, 1990).

Cognitive regulation, which refers to how individuals engage in a recursive process utilizing feedback mechanisms to direct and adjust their learning and problem-solving processes, is crucial for constructing mental model of the complex system to be learned (Azevedo, Guthrie & Seibert, 2004; Manlove, Lazonder & de Jong, 2006), especially when learners are lacking required domain knowledge (Pintrich, 2000). Thus, it is argued that simulation-based inquiry learning environments should include scaffolds to support students' cognitive regulation. This, in turn, would positively support students' inquiry processes starting from hypothesis generation to their interpretation of the findings.

A review of the literature reveals two quite different scaffolding strategies for cognitive regulation that can be ubiquitously embedded into simulation-based inquiry learning environments: (1) question prompts (Ge & Land, 2004); and (2) model progressions (White & Frederiksen, 1990). Providing question prompts to students yielded promising results in some studies (see Ge & Land, 2003); however, others did not yield such positive results (see Greene & Land, 2000; King, 1992). Similarly, model progression did not yield consistent effects on learning performance (Alessi, 1995; Quinn and Alessi, 1994; Rieber and Parmley, 1995). We argue that questions prompts and model progressions may enhance students' cognitive regulation skills, which, in turn, foster their learning of complex system; however, the complexity of the learning domain is a main factor determining the effectiveness of question prompts and model progressions as cognitive regulation scaffolds.

Despite its importance, however, there have been very few systematic studies on which of these strategies are more effective, under which conditions, in scaffolding students' cognitive regulation and learning about complex systems in a simulation-based inquiry-learning environment (Kirschner, Sweller, & Clark, 2006). Moreover, limited number of studies investigated the relationships between cognitive regulation and learning performance in complex systems while existing studies on other technology-based learning environments point out to mixed results.

The purpose of this study is to examine the effect of two promising technology mediated scaffolding strategies, *question prompts* and *model progression*, on ninth-grade students' cognitive regulation skill acquisition and learning about complex ecology systems. We hypothesize that task complexity is an important determinant for whether or not certain cognitive regulation scaffolding strategies are effective. This study seeks to replicate past findings and eliminate some confounding factors.

The reminder of our paper is organized as follows. The next section presents the details of the research study, the research questions, the design of the experiment and the data analysis methods. Then, results are presented. Finally, we conclude with a discussion of the implications of our study and provide suggestions for future research.

## The Present Study

The purpose of this study was to investigate the effects of (1) *question prompts (QP),*and (2) *system dynamics model progressions (MP)*, in scaffolding ninth-grade students' learning of a complex ecology system in a simulation-based inquiry-learning environment. In addition, this study aimed at examining the relationships between cognitive regulation, learning about complex systems, and task complexity. Thus, the following research questions were posed:

1. Does scaffolding with question prompts and system dynamics model progressions affect students' cognitive regulation in the process of developing solutions to problem scenarios ranging in their level of complexity?
2. How do task complexity influence the effectiveness of question prompts and system dynamic model progressions as cognitive regulation scaffolds to support students' learning about complex ecology system?

### Participants and Design

A rural high school in the Midwest of the United States was used as a testbed for this study. 251 ninth-grade biology students were randomly assigned to one of the ten classes. Out of these ten classes, five were randomly assigned to experimental (system dynamics model progression group) condition and five were randomly assigned to control (question prompts group) condition. Of the 219 students, from whom we received both consent and parental assent forms, 113 were in the experimental group and 106 were in the control group. There were 50.6% males and 49.4% females.

### Materials

*Food Chain* is a simulation-based inquiry-learning environment designed to support students while they carry out a sequence of activities that correspond to the steps in the scientific method. *Food Chain* incorporated always-available domain-specific knowledge, scaffolded activities, and analytic tools to help students learn about the processes of experimental design and data analysis, the nature of scientific argument and proof. Scaffolding strategies incorporated in *Food Chain* are aimed at supporting cognitive regulation, metacognitive, and inquiry skills required for implementing scientific methods and effective complex problem-solving. *Food Chain* was originally developed by isee Systems and modified for the purposes of this study.

### Procedure

Participants in both study conditions interacted with the *Food Chain* simulation-based inquiry-learning environment for three weeks. In the first week, participants were introduced to the *Food Chain* simulation-based inquiry-learning system. They were asked to design their own experiments and test their hypotheses to answer a relatively simple complex problem (challenge# 1): which two species (out of eight) can survive in Lake

Mirabile by themselves for 90 days? This first week was intended to familiarize students with the simulation environment and to teach about scientific reasoning processes of hypothesis generation and testing. During this initial run, students in both conditions had access to domain-specific knowledge related to the species in Lake Mirabile, but they did not receive any scaffolding from the system. At the end of each inquiry-cycle, they only received a text-based verification feedback of whether or not their hypotheses were correct.

In the second week, participants tackled a problem scenario with medium complexity (challenge# 2), in which they were asked to identify the smallest number of species that will enable Sunfish, one specie out of the eight species in the simulated environment, to survive for 90 days in Lake Mirabile. Their entire hypotheses, experimental designs, and elaborated reports of the findings were collected.

During the third week, participants tackled a very complex and ill-structured problem scenario (challenge# 3), which called for environmental policy-making. In this problem scenario, students were asked to play the role of an environmental scientist and evaluate the proposal to build 100 new houses on the shoreline at Lake Mirabile from an environmental impact standpoint. All of their hypothesis, results of their experiments, and their elaborated report of recommendations and assessment were collected.

During the second and third weeks, the only difference between the two study conditions was the type of scaffolding strategy provided by *Food Chain* at the end of each inquiry cycle. Following their experiment, participants in the experimental (MP) condition were only provided with a text-based verification feedback (e.g., Nice try…but the species you chose didn't survive! View logic model and try again) and they were given access to the system dynamic model progression, which included an annotated system dynamics model progression (see Figure 1). On the other hand, participants in the control (QP) condition were only provided with a text-based verification feedback and related question-prompts (see Figure 2) to scaffold their new hypothesis generation.

## Measures and Data Analysis

In *Food Chain*, students were required to go through inquiry processes to solve three ecology problems in increasing complexity. For each problem, they were asked to go through the inquiry process at least four times. Each inquiry cycle in *Food Chain* simulation involved the following steps. First, students individually developed a hypothesis using their prior knowledge and information given by the software. Then, they observed the results and analyzed the charts generated by the simulation environment for each variable. Finally, they explained the results. A computer-generated question prompt or an annotated dynamic model progression was provided to the students to interpret their findings. Students were expected to develop their subsequent hypotheses using these scaffolds. Scoring rubrics were used to rate participants' *cognitive regulation* and *complex learning* based on their protocols, which included, for each challenge (complex problem) in the *Food Chain* simulation, four sets of hypotheses, hypotheses justification, results, and explanation of the results.

Cognitive regulation refers to how individuals engage in a recursive process utilizing feedback mechanisms to direct and adjust their learning and problem-solving activities. Therefore, *cognitive regulation* was measured by rating how well the previous inquiry cycle informed the consecutive inquiry cycle, and whether or not the participant responded the cognitive regulation scaffold (question prompt or system model) in generating and justifying the subsequent hypothesis (see Table 1).

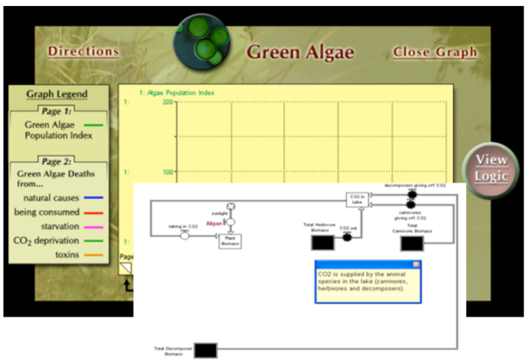



Figure 1. Model progression scaffold received by participants in experiment condition

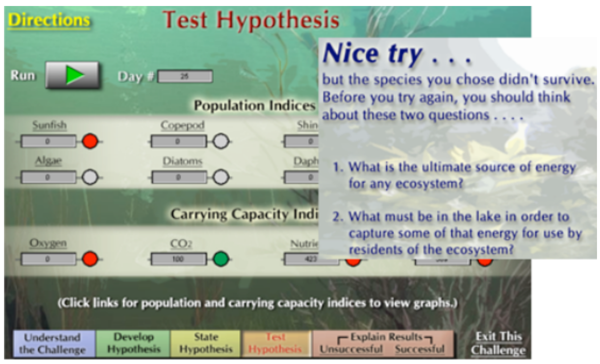



Figure 2. Model progression scaffold received by participants in experiment condition

Table 1: Overview of the scoring rubric for *cognitive regulation* measure

| Coding Category | Number of points given |
|---|---|
| Quality of the hypothesis | 2 points: strong evidences of *using system feedback* to develop the subsequent hypothesis; 1 point: some evidences of *using system feedback* to develop the subsequent hypothesis; 0 point otherwise |
| Quality of justification of the hypothesis | 8 points total: There are two questions per question prompt. 2 points for a clear and correct answer to each question, and 2 points for strong logical justification responding to each question. Deduct 1 point per question if the response is not elaborated clearly. |

In addition, *complex learning* was measured by rating participants' hypothesis, hypothesis justification, and explanation of the results in the inquiry cycles to see if a participant is developing a comprehensive understanding of the lake ecology system and the dynamic interrelationships among its variables. In other words, whether students are able to (1) identify various variables affecting the lake ecosystem (for example, fungi, bacteria, green algae, diatoms, sun energy, $CO_2$, $O_2$, nutrients, etc.), and (2) identify the interrelationships among these variables (for example, diatoms produce the oxygen that fuels bacterial respiratory processes, while bacteria generates the $C0_2$ and nutrients that diatoms require for driving their photosynthetic processes. Both diatoms and bacteria die, of natural causes, feeding the stock of detritus. That stock, in turn, nourishes bacteria that then decompose detritus to create the nutrients that diatoms require for nourishment.) The scoring rubric of the complex learning measured is shown in Table 2.

Two raters coded all protocols based on the analytical rubric system developed by the research team. The interrater reliability was 95% for cognitive regulation measure and 90% for complex learning measure. Any discrepancies of assigned scores were discussed among the raters and the adjudicated score was used. Consequently, a high consensus was reached. Repeated measures ANOVA analysis was conducted to examine change on *cognitive regulation* and *complex learning* measures both after challenge 1, after challenge 2, and after challenge 3 in the *Food Chain* simulation.

## Results

Table 3 summarizes the descriptive statistics for the two dependent variables (cognitive regulation and complex learning) for both control (QP) and experimental (MP) groups along the three challenges: (T1) simpler complex problem where no scaffolding was provided from the system; (T2) more complex problem scaffolded with either QP or MP; and (T3) very complex, environmental policy analysis problem scaffolded with either QP or MP. Below is the statistical analysis report in response to each of the hypotheses tested.

Table 2: Overview of the scoring rubric for *complex learning* measure

| Coding Category | Number of points given |
|---|---|
| Correct variables | 2 points: all correct variables are identified; 1 point: some correct variables are identified; 0 otherwise. |
| Energy cycle | 1 point: evidences of understanding that plants harness energy from the sun; 0 point otherwise |
| $CO_2/O_2$ cycle | 4 points: evidences of understanding that some species are providing $CO_2$ and consuming $O_2$ in the ecosystem and other species are providing $O_2$ and consuming $CO_2$ in the ecosystem; deduct one point each if the students could not identify $CO_2$ , $O_2$, consumption or production cycle. |
| Nutrient cycle | 3 points: evidences of understanding that species are providing for other species for nutritional needs; deduct one point each if (1) the student identify do not have the complete food cycle, (2) cannot identify decomposer role in the system, and (3) do not give a clear explanation. |

Table 3. Descriptive statistics for cognitive regulation and complex learning measures

| *Measures* | Control (QP) Group | | | | | | Experimental (MP) Group | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | | Std. Dev. | | | Mean | | | Std. Dev. | | |
| | T1 | T2 | T3 | T1 | T2 | T3 | T1 | T2 | T3 | T1 | T2 | T3 |
| Cognitive Regulation | 2.01 | 4.21 | 5.72 | 1.03 | 1.13 | 1.95 | 1.97 | 4.15 | 6.20 | 0.96 | 1.70 | 2.16 |
| Complex Learning | 1.84 | 3.45 | 4.33 | 0.89 | 0.49 | 1.16 | 1.64 | 3.75 | 6.92 | 1.04 | 0.89 | 1.30 |

## Effects on Cognitive Regulation

The first research question asked whether scaffolding with question prompts and system dynamics model progressions affect student cognitive regulation in the process of developing solutions to complex problem scenarios.

As evident in the descriptive statistics shown in Table 2, participants in the MP condition had lower mean scores than the participants in the QP condition in *cognitive regulation* after the first challenge in the *Food Chain*, during which no cognitive regulation scaffolding was provided by the system. However, the repeated measures ANOVA analyses did not indicate significant differences between the two groups in these two dependent variables $F$ (2, 209) = 0.47, $p$ > .05, $\eta^2$ = .025. The results indicated that the participants in both conditions were comparably on equal basis after they completed the first challenge.

However, the repeated measures ANOVA analyses revealed a significant main effect in participants' cognitive regulation of inquiry learning in both QP and MP conditions between the first challenge, where no scaffolding were provided by the system, and the second challenge, where participants received either scaffold $F$ (2, 209) = 0.47, $p$ > .05, $\eta^2$ = .025, which supported the hypothesis that students, when received either cognitive regulation scaffold, would perform significantly better than when cognitive regulation was not scaffolded during inquiry learning. Nevertheless, after completing the second challenge in the *Food Chain* simulation-based inquiry-learning environment, the repeated measures ANOVA analyses did not indicate significant differences between the two groups with respect to their cognitive regulation skills.

After the third challenge, which included a very complex environmental policy making problem, the repeated measures ANOVA analyses revealed a significant main effect in favor of the MP group, $F$ (2, 209) = 0.47, $p$ > .05, $\eta^2$ = .025, which supported the hypothesis that, in highly complex learning tasks, providing students with annotated system dynamic model progression was more effective than question prompting as a cognitive regulation scaffold in simulation-based inquiry learning.

## Effects on Complex Learning

The second research question asked whether scaffolding with question prompts and system dynamics model progressions affect students' learning about complex systems, hence, complex problem-solving skill acquisition.

As seen in Table 2, participants in the MP condition had lower mean scores than the participants in the QP condition in *complex learning* after the first challenge in the *Food Chain* simulation-based inquiry-learning environment, during which no cognitive regulation scaffolding was provided by the system. However, the repeated measures ANOVA analyses did not indicate significant differences between the two groups in these two dependent variables $F$ (2, 209) = 0.47, $p$ > .05, $\eta^2$ = .025. The results indicated that the participants in both conditions were comparably on equal basis when they completed the first challenge.

However, the repeated measures ANOVA analyses revealed a significant main effect in participants' cognitive regulation of inquiry learning in both QP and MP conditions between the first challenge, where no scaffolding were provided by the system, and the second challenge, where participants received either scaffold F (2, 209) = 42.66, p < .01, $\eta^2$ = .53, which supported the hypothesis that students, when received either cognitive regulation scaffold, would learn about complex systems better than when cognitive regulation was not scaffolded during inquiry learning. Nevertheless, after completing the second challenge in the *Food Chain* simulation-based inquiry-learning environment, the repeated measures ANOVA analyses did not indicate significant differences between the two groups with respect to complex learning.

For the third challenge, which included a very complex environmental policy making problem, the repeated measures ANOVA analyses revealed a significant main effect in favor of the MP group, F (2, 209) = 49.86, p < .01, $\eta^2$ = .57, which supported the hypothesis that, in highly complex learning tasks, providing students with annotated system dynamic model progression was more effective than question prompting to support students' learning of complex ecology systems during simulation-based inquiry learning.

## Discussion

The purpose of this study was to investigate the effects of two promising cognitive-regulation scaffolding strategies, *question prompts* and *model progression*, on ninth-grade biology students' (1) cognitive regulation of inquiry learning; and (2) learning of a complex system (lake ecosystem) in a simulation-based inquiry-learning environment. This study also addressed the effect of task complexity on the effectiveness of students' cognitive regulation and their successful learning in complex systems.

The results of this study pointed out to a number of important issues. First, the results showed that both scaffolding strategies were effective in providing the cognitive regulation support for students using a simulation-based inquiry-learning system. However, as the complexity of the problem increased, model progression was significantly more effective than question-prompts. This is consistent with earlier studies, which suggested the ineffectiveness of question prompting as a self-regulation scaffold in situations where students have insufficient prior knowledge, which plays an important role in elaborated learning (Ge & Land, 2003; Greene & Land, 2000; King, 1992).

This finding also explains some of the mixed results in earlier studies regarding the effect of model progression on self-regulation. For instance, Quinn and Alessi (1994) performed a study, in which students has access to a simulation of a spread of a disease within a simulation. Students had to manipulate two to four input variables to minimize the value of one of the output variables. Their data revealed that model progression had no overall positive effect on performance. It should be noted, however, that the domain that was used by Quinn and Alessi (1994) was quite simple: the variables in the model did not interact. In another study with simulation of a more complex system (multimeter) Alessi (1995) found that model progression was beneficial for initial learning and for transfer. Similar positive results were observed in Rieber and Parmley's (1995) study in the area of Newtonian motion.

Secondly, the results of this study confirmed the findings from the earlier studies, which suggested using question-prompts as a self-regulation support early in the simulation. For instance, Showalter (1970) recommended using question-prompts early in the simulation to focus learner attention to specific aspects of the simulation. Similar studies by Zietsman and Hewson (1986), Tabak, Smith, Sandoval, and Reiser (1996), White (1984; 1993) also pointed out to the effectiveness of such strategy in different domains for providing self-regulation (especially for planning) support during simulation-based inquiry learning.

Finally, this study provided empirical evidence for the link between cognitive-regulation skills and learning in complex systems (i.e., complex problem-solving), which was suggested in earlier studies (see Butler & Winne, 1995; Davis, 2000; Lin & Lehman, 1999; Schraw, 1998). In this regard, there was no significant difference between the impact of question-prompts or model progression when the problem was a simpler complex problem. However, as the complexity of the problem increased, cognitive scaffolding by model progression had significantly higher impact on learning than cognitive scaffolding by question-prompts.

In future studies, it would be beneficial to investigate the effectiveness of both scaffolding strategies on far-transfer complex problem-solving tasks. These studies should also factor in individual student differences and investigate whether there are differences among low-achieving and high-achieving students. Students who participated in our study had not received any prior instruction on lake ecosystems; the domain-specific knowledge required for inquiry learning was constantly accessible, on demand, in the *Food Chain* simulation environment. Directions included in the system encouraged students to read the domain-specific information prior to generating hypothesis. Therefore, in this study, an attempt was made to bring individual differences to a minimum possible level in order to rule out confounding factors that might have affected the results of this investigation.

Overall, this study provided empirical evidence to the critical importance of embedding appropriate feedback strategies in technology-based learning environments (see Cohen, 1985). This corresponds with the current discussion about the failure of the popular movements of constructivist learning environments adopting discovery, problem-based, experiential, and inquiry-based teaching approaches and provide support to the arguments by Kirschner et al. (2006) and Mayer (2004). We hope to see more studies investigating the effectiveness of technology-mediated scaffolding strategies so that an appropriate design theory for instructional simulations may arise. Current attempts, though interesting, are necessarily fragmentary and incomplete (Thurman, 1993; de Jong & van Joolingen, 1998). Coupled with such a theory, inquiry learning with simulations can fulfill its promise in learning and instruction.

## References

Alessi, S. (2000). Building versus using simulations. In J. M. Spector & T. M. Anderson (Eds.), Integrated and holistic perspectives on learning, instruction, and technology: Understanding complexity (pp. 175-196). Dordrecht: Kluwer Academic Publishers.

Azevedo, R., Guthrie, J. T., & Seibert, D. (2004) The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research*, 30, 87–111.

Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245–281.

Cohen, V. B. (1985). A reexamination of feedback in computer-based instruction: Implications for instructional design. *Educational Technology, 25*(1), 33–37.

Davis, E.A., & Linn, M. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education, 22*(8), 819–837.

De Jong, T. (2006) Scaffolds for computer simulation based scientific discovery learning. In J. Elen & R. E. Clark (Eds.), *Dealing with complexity in learning environments*. London: Elsevier Science Publishers.

de Jong, T. & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179-202.

Dörner, D. (1996). *The logic of failure: Why things go wrong and what can we do to make them right* (R. Kimber & R. Kimber, Trans.). New York: Metropolitan Books.

Dörner, D., & Wearing, A. J. (1995). Complex problem solving: Toward a (computer-simulated) theory. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 65-99). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185-222). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research and Development, 51*(1), 21–38.

Ge, X., & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development, 52*(2), 5-22.

Greene, B. A., & Land, S. M. (2000). A qualitative analysis of scaffolding use in a resource-based learning environment involving with the World Wide Web. *Journal of Educational Computing Research, 23*(2), 151–180.

Greeno, J. G. (1989). Situations, mental models, and generative knowledge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing* (pp. 285-318). Hillsdale, NJ: Erlbaum.

Hogan, K., & Thomas, D. (2001). Cognitive Comparisons of Students' Systems Modeling in Ecology. Journal of Science Education and Technology, 10(4), 319-345.

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology: Research and Development, 45*(1), 65-95.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*(1), 111–126.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41(2), 75-86.*

Lavoie, D.R., & Good, R. (1988). The nature and use of predictions skills in a biological computer simulation. *Journal of Research in Science Teaching*, *25*, 335-360.

Lin, X. & Lehman, J. (1999). Supporting learning of variable control in a computer based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, 36, 837-858.

Linn M. C., Davis E. A., & Bell, P. (2004) Inquiry and technology. In M.C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 3–28). Mahwah: Lawrence Erlbaum Associates.

Manlove, S., Lazonder, A. W., & de Jong, T. (2007). Regulative support for collaborative scientific inquiry learning. *Journal of Computer Assisted Learning*, 22, 87-98.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist,* 59(1), 14-19.

National Research Council (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC.: National Academy Press.

Quinn, J., & Alessi, S. (1994). The effects of simulation complexity and hypothesis generation strategy on learning. *Journal of Research on Computing in Education*, *27*, 75-91.

Pintrich P. (2000) The role of goal orientation in self-regulated learning. M. Boekaerts, P. Pintrich & M. Zeidner (Eds.), In *Handbook of self-regulation* (pp. 451–502). Academic Press, London.

Rieber, L.P., & Parmley, M.W. (1995). To teach or not to teach? Comparing the use of computer-based simulations in deductive versus inductive approaches to learning with adults in science. *Journal of Educational Computing Research*, *14*, 359-374.

Rumelhart, D. E., Smolensky, J. L., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 7-57). Cambridge, MA: The MIT Press.

Sabelli, N. H. (2006). Complexity, technology, science, and education. *Journal of the Learning Sciences*, 15, 5–9.

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences*, *1*, 201-239.

Schraw G. (1998) Promoting general metacognitive awareness. Instructional Science 26, 113–125.

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: Like a bridge over troubled water. *Instructional Science, 29*(4-5), 403-427.

Seel, N. M. (2006). Mental models and complex problem solving. In J. Elen & R. E. Clark (Eds.), Handling complexity in learning environments: Theory and research (pp. 43-66). Elseiver, Ltd.

Showalter, V.M. (1970). Conducting science investigations using computer simulated experiments. *The Science Teacher*, *37*, 46-50.

Shute, V. J., & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments*, *1*, 51-77.

Simmons, P. E., & Lunetta, V. N. (1993). Problem-solving behaviors during a genetics computer simulation: beyond the expert/novice dichotomy. *Journal of Research in Science Teaching*, *30*, 153-173.

Tabak, I., Smith, B. K., Sandoval, W. A., & Reiser, B. J. (1996). Combining general and domain-specific strategic support for biological inquiry. In C. Frasson, G. Gauthier & A. Lesgold (Eds.), *Intelligent Tutoring Systems* (pp. 288-297). Berlin, Germany: Springer-Verlag.

Thurman, R. A. (1993). Instructional simulation from a cognitive psychology viewpoint. *Educational Technology Research & Development*, *41*, 75-89.

Van Joolingen W. R., De Jong T., Lazonder A.W., Savelsbergh, E. & Manlove, S. (2005) Co-Lab: Research and development of an on-line learning environment for collaborative scientific discovery learning. *Computers in Human Behavior*, 21, 671–688.

White, B. Y. (1984). Designing computer games to help physics students understand Newton's laws of motion. *Cognition and Instruction*, *1*, 69-108.

White, B.Y., & Frederiksen, J.R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, *42*, 99-157.

White, B.Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction*, *10*, 1-100.

Wilensky, U., & Reisman, K. (2006). Thinking Like a Wolf, a Sheep, or a Firefly: Learning Biology Through Constructing and Testing Computational Theories-An Embodied Modeling Approach. Cognition & Instruction, 24(2), 171-209.

Young, R. (1983). Surrogates and mappings: Two kinds of conceptual models for interactive devices. In: D. Gentner, & A. L. Stevens (Eds), *Mental models* (pp. 35-52). Hillsdale, NJ: Erlbaum.

Zietsman, A. I., & Hewson, P. W. (1986). Effect of instruction using microcomputers simulations and conceptual change strategies on science learning. *Journal of Research in Science Teaching*, *23*, 27-39.

# Made by Hand: Gestural Practices for the Building of Complex Concepts in Face-to-Face, One-on-One Learning Arrangements

Stephanie Scopelitis, University of Washington, scope@u.washington.edu
Siri Mehus, University of Washington, smehus@u.washington.edu
Reed Stevens, Northwestern University, reed-stevens@northwestern.edu

**Abstract:** Our study investigates how the hands and body can serve as communicative, cognitive and interactional resources when they are employed in the teaching and learning of complex science and math concepts. We consider how speakers and hearers produce, take up, and re-use gestures to support explanation and conceptual formation in expert-to-non-expert conversation. The study suggests that in face-to-face, one-on-one interaction gesturing plays an important role in facilitating the joint achievement of new conceptual understanding for novice learners. We believe that the gestural practices used to explain complex concepts in informal interactions can be profitably extended to the teaching and learning of STEM (science, technology, engineering and mathematics) subjects in other learning arrangements such as the classroom.

## Introduction

Language is an important resource for teaching and learning, but it is not the only one. The intent of this study is to closely examine how both non-verbal and verbal modalities are employed together in face-to-face, one-on-one interactions in which experts in the fields of math and science explain complex concepts to non-experts. Paying attention to how multiple modalities are employed in informal face-to-face, one-on-one explanations of complex concepts can offer valuable insight into how we design practices for teaching and learning in other learning arrangements such as the classroom.

Our study particularly investigates the coordination of gestures and body movement as we as speakers and hearers jointly produce, take up, and re-use gestures to support explanation and concept formation in the context of conversation. We consider gestures and body posturing as participating elements in communication (Becvar, Holland & Hutchins, 2005; Goodwin, 2000; Hutchins, 1996; Kendon, 2001, 2004; Stevens & Hall, 1998; Streeck, 1994, 1996) as we also examine the hands and body as explicit, tangible teaching tools.

The study illustrates how gesture facilitates the joint achievement of new conceptual understanding for novice learners in face-to-face interaction. We believe that the gestural practices used to explain and understand complex concepts in informal interactions can be profitably extended to the teaching and learning of STEM subjects across learning sites. Our study concludes with the suggestion that as we create ever more sophisticated educational designs, we may need to pay more attention to the potential of the most readily available resources – the hands and the body.

## Theoretical Background and Perspective

Our overarching interest is in how knowledge is produced in interaction. We see learning as an interactional process that is dynamically constructed through actions and inter-relationships among resources (Hutchins, 1995; Stevens & Hall, 1998). These resources include participants, settings, multiple modalities, objects and sequences of activities. In our analysis, we employ the concept of *learning arrangements*, which refers to social and material configurations that people arrange for learning, with the question of whether learning is actually achieved being an empirical one for investigation (Stevens, Satwicz & McCarthy, 2007). This concept allows us to examine the resources made available in face-to-face, one-on-one learning arrangements and compare them with other learning arrangements. We are therefore able to treat gesture as both a technical resource and an interactional resource in the active and collaborative enterprise of teaching and learning. We consider how the hands act as media for representations and how the hands and body participate in coordinating available, external representations in interaction to achieve the end task of understanding.

Our study concentrates on teaching and learning in STEM fields (science, technology, engineering and mathematics) and provides perspective on the view that teaching and learning in the STEM fields involves the use of multiple representations such as graphs, models, and charts (Kress, 2001;

Stevens & Hall, 1998). We consider how the hands act as flexible, manipulable, and literally "ready-to-hand" embodied resources for representation. We also respond to recent literature that looks at how professionals in science fields build theory in collaboration. Studies suggest that idea construction and knowledge communication in professional settings happen through an interactional exchange that employs multiple resources that include gestures (Becvar et al, 2005; Hall, Stevens & Torralba, 2003; Ochs, Gonzales & Jacoby, 1996; Stevens & Hall, 1998).

Much of the recent work on the use of gestures in teaching and learning takes a psychological orientation. Such research has suggested that gestures in the classroom can be used to reinforce and extend meaning provided in speech, can aid the gesturer in thought production, and can reveal cognitive processes as a means for evaluating understanding and guiding instruction (McNeill, 1992, 2005; Singer & Goldin-Meadow, 2005; Wagner, Alibali, Flevares & Goldin-Meadow, 1997; Ozyurek, 2002). We expand upon this psychological orientation as we consider not just the relationship between gesture use and cognitive processes, but also the role that gestures play in communication. Attending to both aspects simultaneously provides a more complete and realistic view of how gestures participate in teaching and learning.

## Methods

The study employs the methodology of microanalysis of videotaped interaction, which involves fine-grained, qualitative analysis of sequences of talk and action. This process of analysis allows for the close investigation of moment-to-moment, turn-by-turn, interactional exchanges giving detailed information about what is happening in a particular instance.

The database consists of video-recorded conversations in which experts in STEM fields explained complex concepts to non-experts. Participants came from a pool of volunteers who responded to requests for participation that were distributed via e-mail to STEM departments of a large university. Conversations were held primarily in the offices and labs of the experts and lasted between 30 and 60 minutes. Participants discussed topics of their choosing which they deemed as relevant concepts to their defined discipline and/or research. The recorded interactions were logged and transcribed and the verbal and gestural techniques used to explain concepts were identified and analyzed.

## Data Sources and Evidence

In this paper we analyze segments from three interactions in which experts (two professors and one graduate student) explain concepts in microbiology, mathematics and animal biology in one-on-one conversations with a non-expert. These segments exemplify moments in our data in which the interactions take a conversational format for exchange of knowledge, as opposed to a lecture-like format in which the expert delivers information and the learner is relatively passive. Close examination of the use of gesture in these exchanges shows that the gestures are not merely the outward manifestation of the speaker's cognition, but rather can be co-constructed with a recipient in the service of building conceptual understanding.

## Results

Through our analysis we demonstrate that gestures and posturing are not just the unconscious by-product of a speaker's cognitive processes, but rather often are tactically employed for communicative purposes that respond to particular contingencies of the situation and perceived learning needs of a recipient. As evidence for this perspective we outline examples showing how gestures are (1) built step-by-step for and with the recipient, (2) positioned and adjusted for the recipient, (3) used to correct the learner's misunderstanding (in response to the learner's gesture), (4) used by both the speaker and the hearer to drive the interaction toward the shared goal of achieving understanding, (5) treated as a communicative artifact by speaker and hearer. Finally, we show how, in a single segment, gestures can be seen to support both cognitive and communicative processes.

### Example 1: Constructing a Gestured Object for the Learner

The first segment shows how an expert engages in the process of building, step by step, a gestured object. Sakina is a graduate student in molecular biology. She is explaining the replication of DNA. She first explains the basic architecture of DNA. To do so, she constructs a model layer by layer, enlisting her hands as the material for construction. Sakina begins the construction for the learner/hearer by saying, "Then there are two strands," as she simultaneously lifts her two index fingers, positioned side-by-side, and lowers them--drawing two lines in space (see Figure 1).

Figure 1. "Then there are two strands"

Sakina then raises her left index finger, pointing up, establishing it as a model of the DNA strand. As she says, "One strand we call it the five prime end and the three prime end?" she lifts her right hand and points to a spot toward the tip of her finger and then another spot towards the base of her finger. The right hand thus takes on an indexical role as the left hand models the DNA strand (see Figure 2).

Figure 2. "Five prime end and the three prime end"

Once this basic structure has been established, Sakina alters it to show the learner that the strands go in opposite directions. She does this by straightening her right index finger and then flipping the hand so that the finger is now parallel to the left index finger, but pointing downward. She completes this movement as she utters "anti-parallel" (see Figure 3).

Figure 3. "Anti-parallel"

The gestured object is sequentially constructed for the benefit of the hearer. This is evidenced by 1) the gaze direction of the participants, 2) the speaker's positioning of the gestured-object and 3) the speaker's adjustment of the gesture.

As she begins to set up the model, Sakina shifts her gaze to her hands. The hearer tilts her head to see Sakina's hands (see figure 4). This move by the hearer displays her recognition that the built gestured-object is for her, and that this information is consequential for her understanding.

Figure 4. The hearer tilts her head to see S's gesture; S repositions the gesture.

Sakina immediately responds to the hearer's head tilt by repositioning the gestured-object so the hearer can see it. Sakina's adjustment is a marker that the gestured-object is intentionally constructed and placed for the learner. The speaker's repositioning also works to outline the interactional field. The speaker's body acts as frame, defining a field of action. The speaker and hearer conjointly establish this workspace: first the hearer tilts her head and then the speaker adjusts.

### Example 2: Repairing Understanding through Gesture

In the following example, gestures participate in the repair of a misconception. After Sakina reaches a completion point in her explanation, the learner initiates a turn by putting up two fingers, thus taking up the previously constructed gesture-object. As she does this she says, "Okay so you have the two strands." The gestured-object she displays, however, shows the parallelism of the strands but fails to show the opposing directionality (see Figure 5).

Figure 5. "Okay, so you have the two strands."

Sakina moves to repair the misunderstanding, not through speech but through action. Sakina demonstrates the anti-parallelism with her hand. As she does this, she repeats the learner's previous phrase, "so you have the two strands." The language is the same; it is the gesture that provides the correction. In response, the learner flips her hand to mirror Sakina's and repeats "two strands" (see Figures 6 and 7).

Figure 6. "So you have two strands."

Figure 7. "Two strands."

## Example 3: Assuming the Perspective of the Learner

Another indicator that gestures can be intentional teaching tools is when a speaker shifts position to assume the hearer's perspective on an already built gestured-object.

Max, a professor of geometry, explains the concept of four-dimensionality. First, he explains a more basic concept: three-dimensionality. Max creates the object using the fingers to represent three dimensions (see Figure 8). Once he settles on this construction, he then pauses in the process of explanation. He holds the gesture in its established space and, as though detaching himself from his role of maker of the gestured-object, he steps out of conversational space, crosses the interaction field and observes the built object from the learner/hearer's visual perspective (see figure 9).

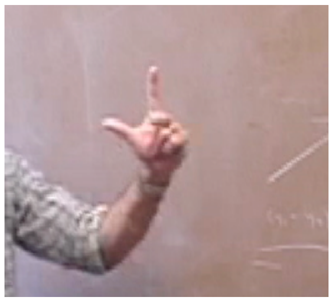


Figure 8. Three dimensions

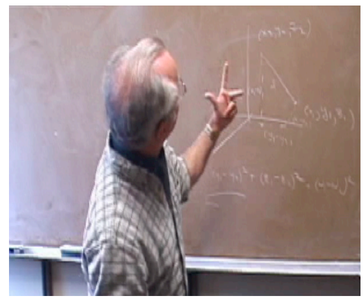


Figure 9. Takes learner/hearer's perspective

Here is a clear embodied representation of recipient design (Sacks, Schegloff & Jefferson, 1974), as the speaker not only constructs and displays the gesture, but steps into the physical position of the learner/hearer in order to "check out" how the gesture looks from that perspective.

## Example 4: Thinking through Gesture Together

A fourth segment shows gesture playing a major role in an extended process in which the expert and learner work toward a shared understanding of a scientific concept. In this case, the expert, Mara, (a biologist) is explaining how fish gills function. Mara, in response to the learner's question about the shape of a fish's gills, moves the discussion away from the process of oxygenation and on to the relationship of surface area to oxygen consumption. In this sequence we see several of the phenomena discussed so far. For instance:

- gestured-objects are constructed for the benefit of the learner;
- gestures are used to repair and clarify the learner's misconceptions; and,
- gestures participate in the moment-to-moment assessment of where the learner is in the process of understanding.

In addition, this segment also shows how the function of the gestured-object may shift as it is passed back and forth by the expert and learner as it also reveals how the body organizes this exchange.

There is a dynamic development of a single gesture that happens as the gesture is passed between the interacting participants. The specific gesture that is the focus of this discussion is made up of an open hand with the thumb pulled to the palm and four fingers loosely separated (see Figure 10). The gesture refers to the gills of a fish. Mara introduces the gestured-object by placing it in the center of the interactional space in the learner's sight line and asking, "But what does this give you when you are like that?" The central placement of the gesture and the use of deictic pronouns "this" and "that" direct the learner's attention to Mara's gesturing hand. The learner, in the turn to follow, takes up the gesture and reproduces it with her own hands (see Figure 10).

Figure 10. "But what does this give you when its like that?"
Speaker displays and hearer mimics "gill" gesture.

The learner's taking up of the gestured-object acts as a demonstration for the participants that the "hand-as-gill" gesture has now been established as a shared object central to the explanation.

When Mara presents the hand-as-gill gesture it is acting as a representational model. When picked up by the learner the gesture shifts to become a manipulable model used to work out an idea, i.e., a thinking tool. This shift of function is evident from the learner's treatment of the gestured-object. She pulls the gesture from the center of the interactional space and to its periphery. She draws it closer to her own body and performs actions on the gestured-object in an abbreviated manner. The gesture, now a thinking tool in the hands of the learner, is used by the learner to explore various relationships between a gill and water. As the learner's left hand maintains the hand-as-gill gesture, the right hand is used to simulate various aspects of the movement of water over the gill such as pressure, impact and flow (see Figure 11). For example, in one simulation the learner moves the fingers of her right hand (the right hand representing water) between the fingers of the left hand (the left hand acting as gill) and utters "it [water] can get through."

Figure 11. Learner uses gestured-object as a thinking tool.

Other uses of the hand-as-gill gestured-object spin out of this interactional activity. In addition to acting as an evaluative tool and a thinking tool, the gesturing also becomes an interactional tool used specifically to organize and *co-coordinate* the teaching and learning. Mara and the learner use their hands and body as mechanisms to steer the direction of the conversation toward the achievement of the shared goal of understanding the concept in play. Mara uses the body positioning to direct the learning toward the kind of investigation that will get to conceptual understanding, and the learner making use of the evaluative benefits of making her thinking visible, uses the hands and body to *check in* and assure that her exploration is on the right track.

As Mara presents the hand-as-gill gestured-object, her body remains on the very outside edge of the interactional space (see figure 10) while the gestured object takes a central position. By remaining on the periphery and foregrounding the built gestured-object, Mara's body positioning further signals for the learner that she is now to step in and take over the learning. The learner responds by doing just that – picking up the gesture-object and exploring it. As she begins to act upon the gestured-object now in her possession, Mara drops her gesturing hand out of the center of the field. The pass has thus been completed. Mara steps out of explanation mode and allows the learner to slip into exploration mode.

The learner's introspective focus, demonstrated not only by the abbreviated manner of her action on the gesturing hand, but also suggested by the slight pulling inward of her body to create her own frame around the gesturing, marks the current function of the gesture as a thinking tool. But it is not just a cognitive tool in use for the learner. It is part of a documentation that is meant to be seen by Mara. The learner's choice to reveal her thinking to Mara through the visibility of the gestures and the audible "figuring out" talk suggests that the learner is seeking on-going guidance from the expert. The learner's vacillating gaze from an inward focus on her hands to outward focus toward Mara also functions to keep the expert "on deck" as a resource while the learner works through learner's questioning. Mara picks up the learner's bid for approbation and/or clarification and shows this as she shifts her body back into the space (see Figure 11) when the learner verbalizes an idea that may move her toward an understanding of the concept. In just a few turns, we witness the intricate, multimodal work participants perform in this interactional process of explaining and understanding.

This final segment shows the importance of adopting both cognitive and interactionist perspectives on the use of gesture. We see that gesture may be employed by the learner and the expert as a (1) communicative tool, (2) a thinking tool to work out and organize knowledge, and (3) an interactional tool to organize the trajectory for learning. The various functions the hands and body can assume are determined and negotiated through a complex interactive exchange between speaker and hearer, learner and expert.

## Conclusion

Our analysis shows that the hands and body do something important in the teaching and learning of complex concepts in face-to-face, one-on-one learning arrangements made up of experts and novices. The gesturing hands can be employed as tools to build a representational object that both the speaker and the hearer act upon in order to achieve a shared understanding of a complex concept. The examples we describe show that gestured-objects are constructed for the benefit of the learner, that gestures participate in moment-to-moment ("formative") assessment in face-to-face, on-on-one interaction, and that gestures are used to repair and clarify misunderstandings. Gestures can be used by both the speaker and the hearer as proposals, cues, and requests that steer the interaction toward the shared goal of achieving understanding. These multiple roles that gestures take in interaction demonstrate that the hands and body are prominent resources for learning in these types of learning arrangements.

Our study also provides evidence that gestures in face-to-face, one-on-one interaction can have both cognitive and communicative functions. Seeing gesture as both cognitive and communicative is an improvement on the dominant cognitive view in that it (1) reveals an important way in which people routinely accomplish teaching and learning in informal talk, and (2) clarifies how both the artful production and the perceptive recognition of gesture is a critical interactional activity achieved and negotiated by all participants.

We believe that by examining the interactional activity that happens in face-to-face, one-on-one learning arrangements between experts and novices we can come to better understand the material and social resources used in the explanation of complex concepts. In all three examples from our study, representations made by the hands and negotiated in interaction become vital aspects of the process of explanation. We invite research that explores ways to extend the benefits of gesture activity for concept learning from face-to-face, one-on-one interaction into the classroom. Certainly, by the very nature of the classroom arrangement, the available uses and functions of gestures are different compared to those present in face-to-face, one-on-one arrangements. Particular gesture activities in co-present interaction may be endogenous to that specific learning arrangement. If so, what certain aspects of gesture use in interaction can be organically and effectively performed within current arrangements for classroom learning? It is, for instance, certainly not possible for teachers to have their gestures respond to each and every learner's displayed understanding but it does seem possible, and worthwhile, for at least some of this to happen.

Our study encourages further research into the profitability and possibilities of extending effective gesture practices that happen in face-to-face, one-on-one interaction into the classroom. We also suggest that investigating gesture use in a variety of learning arrangements will offer more insight into how gestures are shaped by and function within social and material configurations and thus further inform educational design and practices.

## References

Becvar, L. A., Holland, J., & Hutchins, E. (2005). Hands as molecules: Representational gestures used for developing theory in a scientific laboratory. *Semiotica*, 156, 89-112.

Church RB, & Goldin-Meadow S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition,* 23(1), 43-71.Garber, G., Alibali, M., & Goldin-Meadow, S. (1998). Knowledge conveyed in gesture is not tied to the hands. *Child Development,* 69(1), 75-84.

Goldin-Meadow, S., Alibali, M., & Church R.B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, 100(2), 279-97.

Goldin-Meadow, S., Kim. S., & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology,* 91, 720-730.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics,* 32 (10),1489-1522.

Hall, R. Stevens, R. & Torralba, A. (2003). Disrupting representational infrastructure in conversations across disciplines. *Mind, Culture, & Activity*, 9(3), 179-210.

Hutchins, E. (1995 ). *Cognition in the wild*. Cambridge: MIT Press.

Kress, G. R. (2001). *Multimodal Teaching and Learning : The rhetorics of the science classroom.*London; New York: Continuum. (Chapter 2 & 3).

Kendon, A. (2001). Gesture as communication strategy. *Semiotica,* 135(1), 191.

Kendon A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Clarendon lectures in management studies. Oxford; New York: Oxford University Press.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Ochs, E., Gonzales, P., & Jacoby, S. (1996). "When I come down I'm in the domain state" :grammar and graphic representation in the interpretive activity of physicists. In Ochs, E.,Schegloff, E. & Thompson, S. (Eds.) I*nteraction and Grammar* (pp.) Cambridge, Mass.:Cambridge University Press.

Ozyurek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language* 46, 688-704.

Singer M., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, 16(2), 85-9.

Sacks, Harvey, Schegloff, Emanuel A., & Jefferson, Gail (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696-735.

Stevens, R., Satwicz, T., & McCarthy, L. (2007). In game, in room, in world: Reconnecting video game play to the rest of kids' lives. In Salen, K. (Ed.), *Ecology of games: MacArthur Foundation series on digital media and learning*. Cambridge, MA: The MIT Press.

Stevens, R. & Hall, R., (1998). Disciplined perception: Learning to see in technoscience. In Magdalene, L. & Blunk, M. L., (Eds.), *Talking mathematics in school: Studies of teaching and learning*. Cambridge, MA: Cambridge University Press.

Streeck, J. (1994). Gesture as communication II: The audience as co-author. *Research on Language and Social Interaction,* 27(3), 239–267.

Streeck, J. (1996) How to do things with things. *Human Studies*. 19(4) 365.

Wagner, Alibali, M., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology.* 89(1), 183.

# *Micros and Me*: Leveraging home and community practices in formal science instruction

Carrie Tzou, University of Washington Bothell, 18115 Campus Way NE,
Bothell, WA 98011-8246, tzouct@u.washington.edu
Philip Bell, University of Washington, 312 Miller Hall, Box 353600
Seattle, Washington 98195, pbell@u.washington.edu

**Abstract:** Critiques of school science from sociocultural perspectives focus on the narrow scope of the science that is presented to students in school, which in turn constrains how children should engage in scientific sense-making in classrooms. We have designed a seven-week instructional intervention, *Micros and Me*, which attempts to (a) make science more personally consequential to students' lives and cultural memberships, and (b) connect authentic scientific practices deeply with students' repertoires of practice. We report on two iterations of *Micros and Me*, focusing our analysis on two design issues: (1) how we attempted to broaden the definition of what counts as school science through a self-documentation task, and (2) how the self-documentation task problematized the conception of "culture" in each of the classrooms. We end with design implications for the design of science learning environments that take as their starting point deep knowledge of students' repertoires of practice.

## Introduction

Research suggests that wanting to participate in science is influenced partly by whether students identify themselves as people who can or cannot do science (Brickhouse, Lowery, & Shultz, 2000), yet unfortunately science as it is traditionally taught in schools has little to do with the science done by scientists or science as it is practiced in children's everyday lives. In traditional school science, there is usually one "right" answer, and one "right" process to find that answer. Uncertainty and controversy are not part of the discourse of school science textbooks, and laboratory exercises in which students attempt to approximate the "right" answer leave little room for discussion and debate (Bell, 2004a, 2004b).

Critiques of school science from sociocultural perspectives focus on the narrow scope of the science that is presented to students (Warren, Ogonowski, & Pothier, 2003; Eisenhart, Finkel, & Marion, 1996), which constrains how children should engage in scientific sense-making in classrooms (Ballenger & Carpenter, 2004). School science is overwhelmingly Western science (Harding, 1998), with a narrow range of norms of practice (Aikenhead, 1996; McIntyre, Rosebery, & Gonzalez, 2001). Furthermore, these norms of practice may not necessarily be consistent with students' cultural ways of knowing. The literature from sociocultural perspectives tells us that there are problematic ways in which science instruction (and school instruction in general) has attempted to leverage students' repertoires of practice. One way is essentializing students according to their presumed membership to certain cultural groups (Nasir, Rosebery, Warren, & Lee, 2007). Cultural essentialization leads to "treating cultural differences as traits" (Gutierrez & Rogoff, 2003, p.19), unchanging over historical time and belonging to every member of a cultural group. One danger in seeing cultural differences as traits instead of dynamic processes that vary between individuals is the tendency to see these traits as explanations for underachievement, behavior, or other academic problems that apply to all members of a cultural group—the deficit model of cultural differences. Instead, Gutierrez & Rogoff argue that we need to see individual students as having their own experiences and histories that are influenced rather than dictated by their membership in certain cultural groups.

Another, related way in which culture is treated problematically in classrooms is making superficial connections between specific cultural practices (often assumed to be practiced by *all* members of a particular cultural group) and classroom learning. Warren, Ogonowski, and Pothier (2003) cite the example of cooking rice and beans as a way of demonstrating phase transitions. While it is fine to bring such examples in the classroom, we need ways to connect students' culturally-based ways of knowing deeply to scientific meaning making in classrooms. Cultural practices need to be constructed as rich sites of complex reasoning that can be applied to discipline-specific reasoning practices in classrooms.

These critiques underlie the importance of expanding what normally "counts" as science to include children's everyday expertise, or what Gutierrez and Rogoff (2003) call "repertoires of practice". Repertoires of practice refer to the dynamic ways in which people engage in activities based on their background experiences, interests, and cultural memberships. Rather than assuming that certain instructional contexts will connect with students' lives and therefore be motivating contexts for learning, we take the perspective that others (Lee, 2007; Warren, et al, 2001; Moll, 2004) have taken that the design of instruction should begin with deep knowledge of students' repertoires of practice and their corresponding areas of expertise and connecting that in meaningful ways to authentic scientific practices. To this end, we have designed a seven-week instructional intervention, *Micros and Me* (Tzou, Bricker, & Bell, 2007), that attempts to (a) make science more personally consequential to students' lives and cultural memberships, and (b) connect authentic scientific practices deeply with students' repertoires of practice. In this unit, we attempted to elicit and leverage students' repertoires of practice around health in order to motivate the study of microbiology and the connection between microbiology and health. We assume that part of what goes in to building a student's repertoires of practice around health is their membership in certain cultural, family, community, or peer groups.

In this paper, we will describe the context for the science instruction at the study site, briefly describe the design of *Micros and Me*, and present findings from analysis of two iterations of the curriculum, specifically around a self-documentation task (Clark-Ibañez, 2004) in which students were asked to document their health practices and connect those practices to a scientific research project. We end with dilemmas and design implications for applying leveraging students' repertoires of practice in the design of discipline-specific science learning.

### Research Questions

In this design, we asked three questions: (1) how can we *elicit and make visible* students' everyday expertise around health in science instruction? (2) how can we *deeply connect* this expertise to disciplinary practices in science? (3) what are implications for design of science learning environments that start with deep knowledge of students' repertoires of practice and connect disciplinary practices to that knowledge?

## Context for science instruction at Granite Elementary

We conducted this design work in collaboration with three fifth-grade teachers at Granite Elementary, a school located in the southern part of a large, urban city in the Pacific Northwest. The school is marked with significant ethnic and linguistic diversity, with a significant percentage of students on free and reduced lunch, and high achievement on the state standardized tests. This design collaboration is part of an ongoing (5+ year) partnership with the fifth grade teachers at Granite Elementary.

### Design of Micros and Me: an overview

This work is tightly connected to a cross-setting team ethnography of thirteen children's activities in and out of school in order to analyze several conceptual themes: personally consequential biology, argumentation, images of science, diversity (cultural, gender, economic) and uses of technology (cf, Bricker & Bell, 2008). The team has followed the children and their families in school, at their homes, and to a wide range of settings in order to understand the everyday sensemaking practices of children across the settings of their lives. Of the thirteen focal participants in the ethnography, four children were in the fifth grade classes in which the designed curriculum was first enacted. The major goal in the design was to understand how to better architect learning pathways between home and school through:

1. meaningfully incorporating ethnographic findings and protocols into the design of the unit, and
2. eliciting and deeply incorporating students' areas of everyday expertise into the science instruction using ethnographic methods.

*Micros and Me* is a re-design of an existing commercially-available kit called *Microworlds* (NSRC, 2002). The elementary science program in the school district to which Granite belongs is composed entirely of commercially-based kits. Every year, three such kits are taught, and *Microworlds* is one of the three kits taught in fifth grade in the district. The decision to re-design *Microworlds* as opposed to one of the other units in the fifth grade science curriculum was the prevalence in the ethnographic data of chronic health problems such as asthma and allergies. Granite Elementary, and many of the students who attend Granite, live in an area of Seattle that has a very high incidence of hospitalization due to asthma—many times higher than the rest of the county. The area of the city in which Granite is located—Granite Hills—is very industrialized, near ports, railroad tracks, and an airport—all of which are major contributors of pollution. Therefore, we chose to re-work *Microworlds* to help students see the connection between consequential health decisions and practices and scientific inquiry.

## Methods and data

We situate this work under the umbrella of *design-based research* (Collins, 1992, Bell, 2004a), in that we are attempting to understand how to design a personally consequential science curriculum that highlights children's everyday cognition at the same time that it engages children in systematic, discipline-specific practices. The literature is thin on work that designs for students about whom we have thick descriptions from their everyday lives. Seldom do designers have firsthand insight into lives of the children for whom they are designing.

*Micros and Me* was enacted for the first time in the spring of 2007 and every year since in all three fifth grade classrooms at Granite Elementary. A variety of data sources were collected during this effort: daily, videotaped observations of every day of the enactment, artifacts the students produced (including science notebooks, final reports and projects, pre/post assessments, and surveys of their everyday activities), and interviews with focus groups of students after the enactment. For the purposes of this paper, we draw on videotaped observations of students and teachers during the self-documentation lessons and the pictures and journals from 34 students across the three fifth grade classrooms. Videotaped observations were content logged and key episodes were transcribed in more detail.

Journals and classroom talk were analyzed for evidence of repeated activity systems that pertained to health and explanations for those activity systems that were rooted in family or cultural belief systems. We were especially interested in understanding whether the self-documentation technique would give us enough insight into students' everyday expertise around health that we could connect that expertise to the classroom scientific inquiry. Therefore, the following analysis is not so much concerned with the particulars of what students took pictures of, but of the potential for eliciting and making visible students' health practices in the context of classroom instruction. We use as a theoretical basis for this analysis the Everyday Expertise framework (Bell, et al, 2006), that highlights three dimensions of analysis: individual, social, and cultural, as well as the accompanying learning phenomena for each. *Cognitive ecology* refers to the diverse terrain of influences on the way an individual cognitively makes sense of, or assigns meaning to, the world. For example, an individual may have certain ideas about herself as a learner (epistemological knowledge) based on cultural, gender, age, and class identities, and these ideas influence the way she acts upon the world (diSessa [2002] for his notion of conceptual ecology). Although these histories and experiences are socially situated, they are considered here in relation to how they influence the perceptions and actions of individuals as they engage in social activity.

*Situated activity systems* refer to activity that takes place within the boundaries of a (socially constructed) place (Goffman, 1961). Individuals can "boundary cross" (Phelan, Davidson, & Cao, 1991, p. 227) as they move between activity systems, each of which contains culturally-based narratives. The social plane is the space where individuals enact their personal epistemologies in social interactions and in situated activity systems. As individuals enter places, they interact not only with each other but also with the cultural narratives and resources available in those places. Cultural toolkits (Swidler, 1986) refer to the ideologies, technologies, and narratives that are uniquely present at a given historical moment. We consider cultural activities and toolkits broadly, to include resources from environmental education, environmentalism, school, ethnic group memberships, and religious participation. These toolkits then shape activities and interact with an individual's own cognitive ecology.

Taken together, these dimensions help us see the practices that students self-document and discuss in class as situated within activity systems in specific socio-historical contexts. For example, when students self-documented a narrow range of health practices relative to the variation that we knew existed based on our ethnographic work, this framework allows us to interpret those findings in light of norms within the situated activity system of "doing school science" that typically allow for only Western science practices to count as school science (Harding, 1998).

## Findings and design implications

In this section we present findings from an initial analysis of the self-documentation task and design implications from those findings for the next design iteration of *Micros and Me*. We present our findings by telling the design story of the curriculum over the first two iterations, which occurred in the spring of 2007 and the spring of 2008.

### Iteration 1: Making visible students' everyday expertise around health through self-documentation

We knew, from our ethnographic findings, that students in this school population engaged in rich and varied repertoires of practice around health. Reeve (in preparation), using Chrisman & Kleinman's (1983) framework of multiple domains of health care, reported that families in the ethnographic study utilized multiple domains (what Chrisman & Kleinman call professional, popular, and folk) to manage their health care. This means that they went outside of established Western medicine in the maintenance and treatment of health-related issues. In this population of students, therefore, we were expecting to see similar varieties of health-related practices. Our first attempt to elicit and make visible students' everyday expertise around health was through student self-documentation. This was a photo-elicitation activity (Clark-Ibañez, 2004) in which students took pictures of practices they engaged in around health and explained how those pictures related to health. We loaned each student a digital camera for one night and asked them to take pictures of activities that that they engaged in around health. We also gave students a worksheet to document and explain the pictures they took. The intention was to gain a deeper understanding of students' repertoires of practice around health in order to connect those practices with the scientific inquiry in the classroom. We hoped that one of the activities documented during the task would be used as the basis for students' research projects at the end of the unit.

After we collected their pictures and their worksheets, we constructed what we called "maps" of their everyday health expertise using a commercially-available comic-strip making software program. These maps showed a subset of the pictures each student took (chosen by the researchers as the practices that had the most potential to be research topics) and the explanations of those practices in the students' own words.

### Construction of self-documentation activity as "school science" activity

In the first iteration of *Micros and Me*, there was wide variation in what the students documented as related to staying healthy and preventing illness. However, most of these fell squarely into what we would categorize as "mainstream" practices, well-documented in Western science as linked to health maintenance (drinking water, handwashing, eating vitamins, etc). We were at first disappointed in these results, since from the ethnographic data we had conjectured that this population of youth would engage in more varied practices around health. However, we soon realized that the self-documentation task was capturing only a partial list of health practices in which students engaged. When teachers conducted discussions with students about their self-documented pictures, we found that the classroom discussions showed potential to be rich sites for further exploring students' repertoires of practice around health. Consider the following brief discussion in one teacher's classroom:

1 Ms. E: One thing I'm always wondering about I think would be my topic is in my culture, in my family, everybody says I'll make you chicken soup, it'll make you feel good. And I always wonder is that just something to make people feel good because it's nice and warm and shows that people care about them, or is there something in the chicken soup that helps you get healthy. So that's a cultural thing that I always wonder about that has to do with being healthy. I want you also to think about tonight is there anything in your culture that has to do with um staying healthy or that people say. Some cultures have special things they wear, or put on their bodies to prevent them from getting sick or help them get healthy. Robbie
2 Robbie: You get a coin, and get this white or green ointment and put it on your back.
3 Ms. E: but is to help you get better or prevent you from getting sick?
4 Robbie: better
5 Ms. E: Ok so that's another cultural thing. Danielle
6 Danielle: there's this dance that you do
7 Ms. E: Ok so a dance. And is it supposed to help you get well?
8 Danielle: Yeah
9 Ms. E: Ok. So there's lots of things that we do in our culture, ethnicities that have to do with getting well. And I still have to find out if they always say I'll make you chicken soup is it going to help you feel better or is there something inside. So tonight as you're thinking about this, I want you to start thinking is there something in your pictures that you would like to research about staying healthy. *(Transcript of Classroom observation, 2007-05-29)*

In the above example, Ms. E was able to elicit culturally-based health practices that had not been documented in Robbie or Danielle's self-documentation tasks. By modeling the cultural practice of making chicken soup, she helps students understand what "cultural practice" means in this context. This is evidenced by Robbie's and Danielle's responses. However, she also says, "I always wonder is that just something to make people feel good because it's nice and warm and shows that people care about them, or is there something in the chicken soup that helps you get healthy", thus indicating a way in which a cultural practice could be leveraged in the context of a scientific research project. She not only asks questions that would allow her to examine that cultural practice scientifically, but also indicates ways in which it could be connected specifically to microbiology and health.

The contrast between the practices documented in the self-documentation activity and the practices students shared during the class discussion are evidence that the students were constructing this activity as a "school science" activity—where a narrow range of practices usually "counts" as science. When the teachers explicitly inquired into practices that were *cultural*, a different range of practices emerged. In the second iteration, we decided to more explicitly address "cultural practices" in the self-documentation task, building in more formal structures to do so.

## Iteration 2--Explicit conversations around cultural practices: problematizing "culture" and the Costco incident

In the next iteration of *Micros and Me*, we had the teachers make two substantial changes to the self-documentation task. First, they presented their *own* self-documentation maps when they introduced the activity, giving students a chance to see their teachers' repertoires of practice around health. We hoped this would prompt the students to think about the self-documentation task more broadly, outside of the narrow confines of Western school science. The second change was instituting a whole-class discussion in which teachers specifically addressed practices as "cultural", in order to get a better idea of what Gutierrez & Rogoff (2003) call the "historied and varied practices" that students engage in around health. What emerged was both a broadening of "what counts" as science and a problematizing of what it means to talk about practices as "cultural".

The whole-class discussion around cultural practices was preceded by a discussion in which teachers simply prompted students to share some of the practices that they had documented. In this discussion, the practices that students chose to share all fell squarely within the confines of Western science: eating off of clean dishes, using Clorox, drinking milk, washing hands, taking vitamins. In the *next* discussion, however, when the teachers addressed practices explicitly as *cultural,* a very different list of practices emerged (see Table 2). We argue that Discussion 2 was able to somehow elicit a qualitatively different list of practices, most of which were documented on the self-documentation maps, but not shared in Discussion 1. Many school science activities would have stopped at Discussion 1, but would have missed the rich variation in practices that were elicited in Discussion 2. Because the space was opened up in Discussion 2 to accept a broader range of acceptable practices in the context of science class, students re-constructed the task as something different from a school science task (as illustrated in the difference between the two lists).

Table 2: Comparison of elicited practices in Discussions 1 and 2 in the second iteration of *Micros and Me*

| Discussion 1 | Discussion 2 |
|---|---|
| • Eating off of clean dishes<br>• Taking showers<br>• Brushing teeth<br>• Drinking milk<br>• Using a vacuum<br>• Washing hands<br>• Taking vitamins<br>• Drinking water<br>• Exercise<br>• Using Clorox | • Mangosteen for allergies<br>• Pho: prevents illness<br>• Tea (black, cinnamon, brown)<br>• Green oil on back, scratching with a spoon: tells you how sick you are and helps you get better<br>• Ginseng: clear face, helps make you feel better |

In the following example, we show a short excerpt from the beginning of Discussion 2 in Ms. Evans's class to examine the pedagogical moves made during this discussion.

Ms. E: Allison and Kristin, because you both know it. How many others of you know what Mangosteen is?
Student: Can you show us an example?
Ms. E: I don't have one, I've never seen it, I know nothing about it. Can you t--So yes, I can. Ok? (holds up a student's self-doc with it)
Ms. E: This is mangosteen. Tell us what you do with it.
Allison: um, when you buy it I put it in a tablespoon and then I drink it.
Ms. E: and so what does m--what does mangosteen--why do you drink it? Does it taste good?
Alice: It gives you energy and then it keeps you from sneezing and allergies.
Ms. E: And so you just take one spoon and it, and it keeps you...
Alice: from sneezing and allergies.
T: Ok, so it keeps you from sneezing and allergies.
Ms. E: Amy, you say you use it as well? Do you use it for the same thing?
Amy: yeah
T: And do you take it when you're sick or do you take it to prevent sickness?
Amy: Every day
Ms. E: So you think it would prevent this.
Ms. E: Do you use it as well? Do you use it for the same thing?
Ms. E: Anybody else use Mangosteen?
Ms. E: So if you look at the jar, and if we look at the jar (gets a student's self-doc), I look at the jar and I've never had Mangosteen, or used it, um, is this something that you think is part of your culture?

Allison: No, you can just buy it at Costco.
Ms. E: You can buy it at Costco. I'm guessing that it's something that your parents or grandparents knew about because as I look around the room, many of us from your culture don't know about it but all the people that do know about it are of the Asian cultures. Does that make sense? And so will all who use it go home and talk to your parents and find out how they knew to use Mangosteen.

In this example, Ms. E makes some pedagogical moves to scaffold the discussion around cultural practices. The first move is a *positioning* move (Harre, et al, 1999) in which Ms. E positions herself, in lines 4 and 21, as someone who knows nothing about the cultural practice in question, drinking Mangosteen. Simultaneously, she positions the two students who documented this practice, Allison and Amy, as experts in this practice, thereby highlighting the value of varied repertoires of practice and in this discussion. Another move Ms. E makes in this discussion is to not assume that all students who might use this product would use it for the same purpose. She asks, in lines 13 and 20, "do you use it for the same thing?", thereby avoiding essentializing the practice to one purpose or one cultural group. We note that this pedagogical move represents a sharp diversion from the "traits" model of cultural practices (see Introduction) towards a model of *practice*—to "bring the multiple dimensions of students' lived experiences to life in the classroom" (Rosebery, McIntyre, & Gonzalez, 2001).

## Defining "culture"

Across all three classrooms, varied definitions of "culture" emerged through this discussion. In the above example, when Allison says that Mangosteen is *not* part of her culture because "you can just buy it at Costco", it brings up the question for us of what it means to talk about culture with students, how one defines it, and how one leverages "cultural" practices in the context of science instruction. In this section we briefly present the varied definitions of culture that emerged through the discussions around the self-documentation activity. We raise this point not because it is surprising that "culture" was a contested term in all three classrooms, but to raise the question (still unanswered for us) of what are design implications for making cultural practices an explicit object of inquiry in science classrooms.

In the following examples, we show classroom excerpts from all three teachers to illustrate the varied definitions of cultures from each teacher:

1. Ms. Love: culture is something that would be different from what I do: *"I also want to be able to pull out what are some of the things that are cultural to us? Anyone know what I mean by cultural to us? So if I walked into your home, what would be different from what would be in my home? If I walked in there sick, and I'm just like (starts coughing, holding her side, acting sick), what are the kinds of things would your parents do for me if they took me in, those kinds of things? Or if you walked into my home sick, what are the kinds of things I would do for you?"*
2. Ms. Evans: culture is based in your family and your ethnicity: "*I'm guessing that it's something that your parents or grandparents knew about because as I look around the room, many of us from your culture don't know about it but all the people that do know about it are of the Asian cultures. Does that make sense? And so will all who use it go home and talk to your parents and find out how they knew to use Mangosteen."*
3. Ms. Williams: culture is handed down from your families: "*So what we're really interested in here is what your family does based on the culture that you come from. Ok? So this is something that comes down for you from your family, from your culture, that's especially important. So not something that you learned in school, but something that you learned from your family of ways to keep you healthy, ways to keep you from getting sick…So raise your hand if something that is on your self-doc, something that's on your map in front of you right now with the pictures on it, or something that you know of off the top of your head, that's something that you do when you get sick or something that you do everyday that's something that your grandparent does, or that you've learned from your family."*

In Ms. Love's example, her definition of culture seems to assume that there are no *shared* cultural practices—that a person's cultural practices only count as "cultural" if they are different than someone else's. In Ms. Evans's example, culture is handed down through family and membership in an ethnic group. In Ms. Williams's case, culture is only something that is learned from your family, *not* something learned in school. While each of these conceptions of culture may be incomplete in some way (ie, excluding peer groups as a site of cultural membership, thinking about shared cultural practices), they represent a starting point for us to think about how to support conversations around culture in classrooms—especially as they pertain to science. For example, the self-documentation gave teachers a glimpse of sources of certain knowledge systems around health ("my dad says that drinking tea gives you more blood"). With the recognition that students' beliefs around health come from strong family histories of knowledge being passed down from one generation to the next (Ms. Williams's definition of culture), teachers are less likely to see these beliefs as "misconceptions" and more as rational ideas. By recognizing (and celebrating) differences between individuals' cultural practices (Ms. Love's definition of culture), we can begin to broaden our definitions of what counts as school science. Finally, in recognizing that some cultural practices are common among people who share certain ethnicities—and by pointing that out in science class, we begin to position students as both experts *and* members of certain ethnic groups, thereby helping students to "affirm their cultural identities" (Ladson-Billings, 1995, p. 6) at the same time that they participate in meaningful scientific learning.

## Implications

This design effort represents our first attempt to incorporate ethnographic data and methods into the design of formal science instruction. In attempting to surface and leverage students' out of school expertise around health, the self-documentation technique shows promise as a way for teachers to gain some insight into their students' out of school practices. Because of the time-consuming nature of ethnographic work, thick descriptions (Geertz, 1973) are formed about a small number of participants. Self-documentation might be a way to help practitioners, with 20 or more students in their classes, gain some insight into who their students are—the multiple and varied repertoires of practice that they engage in and the corresponding areas of expertise they bring with them to the classroom.

This analysis also raises some dilemmas for us about *how best* to leverage students' culturally-based health practices in the learning of discipline-specific practices. Gutierrez and Rogoff (2003) argue that "An important feature of focusing on repertoires is encouraging people to develop dexterity in determining which approach from their repertoire is appropriate under which circumstances" (p.4). While the self-documentation task itself was successful in eliciting students' repertoires of practice around health, more work needs to be done in the design of science learning environments to leverage those practices (while knowing the histories behind them) and support students in understanding when and how to navigate between their various repertoires of practice and where there is overlap or tension between them.

Another implication we see from this study is the challenge of leveraging culturally-based practices *while at the same time* supporting the learning of disciplinary practices and content. At the same time that the self-documentation activity gave teachers insight into how to connect science to their students' everyday lives, we have more work to do to connect these practices with deep disciplinary content.

## References

Aikenhead, G.S. (1996). Science education: Border crossing into the subculture of science. *Studies in Science Education, 27*, 1-52.

Ballenger, C. & Carpenter, M. (2004). The puzzling child: challenging assumptions about participation and meaning in talking science. *Language arts, 81*(4), 303-311.

Bell, P., Bricker, L. A., Lee, T. R., Reeve, S., & Zimmerman, H. T. (2006). Understanding the cultural foundations of children's biological knowledge: Insights from everyday cognition research. In S. A. Barab, K. E. Hay & D. Hickey (Eds.), *Proceedings of the Seventh International Conference of the Learning Sciences (ICLS)* (pp. 1029-1035). Mahwah, NJ: Erlbaum.

Bell, P. (2004a). On the theoretical breadth of design-based research in education. *Educational Psychologist, 39*(4), 243-253.

Bell, P. (2004b). The educational opportunities of contemporary controversies in science. In M. C. Linn, E. A. Davis & P. Bell (Eds.), *Internet environments for science education* (pp. 233-260). Mahwah, NJ: Erlbaum.

Bricker, L. A., & Bell, P. (2008). Mapping the learning pathways and processes associated with the development of expertise and learner identities. In P. A. Kirschner, J. van Merriënboer & T. de Jong (Eds.), *Proceedings of the Eighth International Conference of the Learning Sciences* (pp. 3-206–3-123). International Society of the Learning Sciences, Inc.

Brickhouse, N. W., Lowery, P., & Schultz, K. (2000). What kind of girl does science? *Journal of Research in Science Teaching, 37*(5), 441-448.

Brown, A. L., & Campione, J.C.(1998). Designing a community of young learners: Theoretical and practical lessons. In N. M. Lambert & B. L. McCombs (Eds.), *How students learn: Reforming schools through learner-centered education* (pp. 153–186). Washington, DC: American Psychological Association.

Chrisman, N.J. & Kleinman, A (1983). Popular health care, social networks, an cultural meanings: The orientation of medical anthropology. In D. Mechanic (Ed.), *Handbook of health, health care, and the health professions* (pp. 569-590). New York: The Free Press.

Clark-Ibañez, M. (2004). Framing the social world with photo-elicitation interviews. *The Behavioral Scientist, 47*(12), 1507-1527.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Belknap Press.

Collins, A.(1999).The changing infrastructure of education research: Issues in educational research. In E. C.

Lagemann & L. S. Shulman (Eds.), *Issues in education research: Problems and possibilities* (pp.289–298). San Francisco: Jossey-Bass.
diSessa, A. (2002). Why "conceptual ecology" is a good idea. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 29-60). Dortrecht: Kluwer.

Eisenhart, M., Finkel, E., & Marion, S.F. (1996). Creating the conditions for scientific literacy: A re-examination. *American Educational research Journal, 33*(2), 261-295.
Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.
Goffman, E. (1961). *Encounters: Two studies in the sociology of interaction*. Indianapolis, IN: Bobbs-Merrill.
Gutiérrez, K., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 22*(5), 19-25.
Harding, S. (1998). *Is science multicultural?*: Postcolonialisms, feminisms, and epistemologies. Bloomington, IN: Indiana University Press.
Harré, R. & van Langenhove, L (1999). *Positioning theory: Moral contexts of intentional action*. Malden, MA: Blackwell.
Lee, C.D. (2007). *Culture, literacy, and learning: Taking bloom in the midst of the whirlwind.* New York: Teachers College Press.
Lee, C.D. (1995). A culturally based cognitive apprenticeship: Teaching African American high school students' skills in literary interpretation. *Reading Research Quarterly, 30*(4), 608-631.
Lee, C.D., Spencer, M.B., & Harpalini, V. (2003). "Every shut eye ain't sleep": Studying how people live culturally. *Educational Researcher, 32*(5), 6-13.
McIntyre, E., Rosebery, A., & Gonzalez, N. (2001). *Classroom diversity: Connecting curriculum to students' lives.* Portsmouth, NH: Heinemann.
Moll, L.C., & González, N. (2004). Engaging life: A funds of knowledge approach to multicultural education. In J. Banks & C. McGee Banks (Eds.), *Handbook of research on multicultural education* (Second edition) (p. 699-715). New York: Jossey-Bass.
Nasir, N. S., Rosebery, A.S., & Lee, C.D. (2007). Learning as a cultural process: Achieving equity through diversity. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 489-504). New York: Cambridge University Press.
National Science Resource Center (2002). *Microworlds*. Burlington, NC: Carolina Biological Supply Company.
Phelan, P., Davidson, A. L., & Cao, H. T.(1991) Students' multiple worlds: Negotiating the boundaries of family, peer, and school cultures. *Anthropology & Education Quarterly, 22*(3) 224-250.
Reeve, S., Bricker, L.A., & Bell, P. (in preparation). Cultural influences on youth prior knowledge about health and health care: Implications for science education.
Stevens, R.(2000). Divisions of labor in school and in the workplace: Comparing computer and paper-supported activities across settings. *Journal of the Learning Sciences,9*(4), 373–401.
Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review, 51*(2), 273-286.
Tzou, C., Bricker, L. A., & Bell, P. (2007). Micros & Me: A Fifth-Grade Science Exploration into Personally and Culturally Consequential Microbiology. Seattle, WA: Everyday Science & Technology Group, University of Washington.
Warren, B., Ballenger, C., Ogonowski, M., Rosebery, S., Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of research in science teaching, 38*(5), 529-552.
Warren, B., Ogonowski, M., & Pothier, S. (2003). "Everyday" and "Scientific": Re- thinking dichotomies in modes of thinking in science learning. In R. Nemirovsky, A. Rosebery, J. Solomon, & B. Warren (Eds.), *Everyday Matters in Mathematics and science education: Studies of complex classroom events* (pp. 119-152). Mahwah, NJ: Lawrence Erlbaum.
White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*, 3–11.

# Playing with Food: Moving from Interests and Goals into Scientifically Meaningful Experiences

Tamara L. Clegg, Christina M. Gardner, Janet L. Kolodner
Georgia Institute of Technology, 85 5th street, NW Atlanta GA 30318
tlclegg@cc.gatech.edu, cmgardne@cc.gatech.edu, jlk@cc.gatech.edu

**Abstract:** As science educators, we want all learners to see the relevance of science to their lives and the world in which they live. Achieving this goal, however, has proven to be a difficult endeavor. Many learners see science as useful only in school, and they face difficulties connecting science to the real world and to their own interests and goals. In our research project, Kitchen Science Investigators, we aim to start with learners' interests and goals in cooking. We then help them connect cooking to science, using play to help them see food as an object of investigation. We then transition learners into engaging in authentic scientific practices. In this paper we present three cases that highlight scientifically meaningful experiences for KSI learners and the ways play, facilitation, and artifacts bridge the gap between their interests and scientific practices.

## Introduction

As science educators, we want all learners to see the relevance of science to their lives and the world in which they live. Indeed, this is also a goal set forth by science education standards (Rutherford & Ahlgren, 1991). Achieving this goal, however, has proven to be a difficult endeavor. Many learners see science as useful only in school, and they face difficulties connecting science to the real world and to their own interests and goals (Chinn & Malhotra, 2001). Subsequently, many students are turned off by science. We therefore aim to understand how we can help learners see the relevance of science to their lives and enthusiastically engage in scientific practices. Success at answering this question is essential if we want more citizens to engage in public discussions about global climate change, stem cells, and other scientifically-relevant endeavors.

Researchers and designers of informal learning environments have found that starting with learners' own interests and goals presents opportunities for helping learners to engage in and to see the relevance of scientific, mathematical, and professional practices to their lives. Seiler (2001) created a science lunch group for African American boys in an underserved and underperforming high school around drumming and sports. For those interested in drumming, their discussions and investigations focused on topics such as sound frequency. For those interested in sports the conversations focused on laws of motion in physics. Nasir (2002) studied African American males' use of advanced mathematical understanding in the context of playing basketball to calculate their shooting statistics and to compare themselves to professional basketball players. Shaffer & Gee (2007) leveraged children's interest in video games to engage them in epistemic thinking and reasoning particular to a professional practice. They found that after playing one such epistemic game that required learners to act as urban planners, learners began to think of their own worlds in new ways. They also found that these learners thought in more complex ways about city planning. In each of these approaches, learners' interests and hobbies are engaged to help them move toward learning the "rules of the game" in some discipline or field they might join someday (Shaffer & Gee, 2007).

We, too, have created an approach to engaging more learners in science practices. In Kitchen Science Investigators (KSI), participants learn science and engage in scientific practice as they are creating and perfecting recipes (Clegg, Gardner, Williams, & Kolodner, 2006). Our afterschool and summer camp cooking and science program provides opportunities for learners, who may or may not be interested in science, to learn science through cooking. Our goal was to start with learners' interests and goals in cooking, help them to see food as an object of investigation through play, and then transition them into engagement in authentic scientific practices. We have seen this approach lead to scientifically meaningful experiences for our learners.

In this paper, we explore three experiences KSI participants indicated were scientifically meaningful to them. We seek to find out (1) What is a meaningful scientific experience to a middle-schooler who may not be particularly interested in science? (2)What scientific practices and accomplishments are supported by these scientifically meaningful experiences? (3) How can the having of such meaningful scientific experiences be promoted?

Of course, we cannot begin to answer these questions without defining some terms. To us, a *scientifically meaningful experience* is one in which learners derive meaning relevant to their lives from acting and thinking as scientists. In KSI, they do this through answering a question of interest to them, answering a question useful to themselves or to others, and through accomplishing a valued goal. Sometimes, they recognize the meaningfulness of the experience while it is happening; other times they don't recognize the meaningfulness until some time later when they are in a situation where what they have learned is valuable to someone else. We define *play* as voluntary engagement in recreational activities and less formal practices guided by learners' personal interests, curiosity, or amusement. Play is fun for participants, and during play, participants tend to be guided by trial and error. Science is more systematic than play, but play is often a good way to promote noticing and perspective taking. We find that we often encourage participants in KSI to play with their food so that they will notice attributes that they don't normally think about when cooking or eating. Then we help them move from playful participation to more scientific participation; we do that by helping them experience the value of systematically investigating attributes of food that they discovered during playful moments.

## Background

Science education literature stresses the importance of going beyond teaching learners *about* science, but also helping them to see *how* science is done (Osborne, Collins, Ratcliffe, Millar, & Duschl, 2001). In order to do this, we must get learners engaged in scientific inquiry by helping them to carry out scientific practices the way that scientists do. However, learners often do not engage in the scientific practices of professional scientists because classroom science emphasizes simple experimentation (Chinn & Malhotra, 2001). Helping them to engage in authentic scientific practice in the context of their own goals, we believe will lead learners to have scientifically meaningful experiences of their own. A scientifically meaningful experience has two components. First, a scientifically meaningful experience will involve learners engaging in scientific practice. Second, it is an experience that has personal meaning to learners themselves.

Scientific practices are those actions and pursuits relevant to scientific reasoning to test and explain phenomena. Such actions include, but are not limited to generating research questions, designing experiments to answer questions, controlling variables, making hypotheses and predictions, making observations, taking measurements, developing theories, and studying others' research (Chinn & Malhotra, 2001; Osborne et. al, 2001). While these actions may be encouraged in traditional classrooms, they are typically enacted in experimentation that is simple and fixed which is different from the experimentation of scientists. In contrast to simple science typically done in classrooms, *authentic scientific practice* that scientists engage in involves doing science (1) in the context of real-world problems, (2) where the full range of variables can be tested and the full range of outcomes may be unknown, and (3) where procedures for answering questions are chosen at least partially by participants, rather than being rigidly prescribed and ordered (Chinn & Malhotra, 2001; Gleason & Schauble, 1999a, 1999b). Authentic scientific practice thus includes, among other things, designing experiments and investigations to answer societal or real world questions, establishing rigor in these pursuits, and checking for the mistakes and misinterpretations of one's own work and the work of others.

A scientifically meaningful experience will not only involve scientific practice, but it will also be a personally meaningful experience for learners. Learners have much richer learning experiences when they can not only learn from the experience but when it also connects to their own interests, passions, and experiences (Resnick, Bruckman, & Martin, 1996; Shaffer & Resnick, 1999). We recognize personally meaningful experiences when learners use what they learned from the experience later and when they report deriving meaning for their own particular goals or interests from those experiences.

Based on our experiences, personally meaningful experiences for middle-schoolers often involve two forms of play: social play and physical manipulation. Talking, joking, laughing, and physical play that often occurs between friends and reinforces the social bonds between peers characterize social play. Play can also refer to tinkering that learners do with objects in the environment -- physically manipulating objects by touching, feeling, smelling, and/or pressing them. With respect to scientific participation, we have found some learners prefer designing experiments that specify actions to be done to objects in their scientific pursuits prior to beginning experimentation. Others prefer to design complex procedures that address problems that arise along the way by trial-and-error, tinkering with objects (Clegg & Kolodner, 2007). Both types of participation fit within the pursuit of science, but the latter is facilitated by playing with objects and can be harder to distinguish as scientific practice, especially when they are highly engaged in social play.

A scientifically meaningful experience will therefore connect a learners' scientific pursuit to their own interests, goals, and style of participation. Our goal then as designers and educators is to help learners participate in science in a manner that is authentic to the disciplines of science. In addition, we want to help them see scientific practices and their understanding gained as a result of engaging in those practices as relevant to their everyday lives, their interests, and their passions. Constructionist learning environments have been successful at designing tools that enable these types of meaningful experiences. Science and Math education literature has shown that these meaningful experiences can indeed happen in natural settings (Crowley & Jacobs, 2002; Nasir, 2002). Our work seeks to inform how we can design activities and facilitation support that enable these experiences in informal learning environments.

## Design of KSI

Our goal in designing KSI was to engage learners in scientific practices through providing them with hands-on cooking and science experiences in an informal after school setting. KSI is informal, in several ways:

(1) participation in the program is not a requirement, rather, on a weekly basis participants have the choice to decide if they want to participate in the program and (2) there are no formal science or cooking standards we are trying to meet, thus (a) participants are able to choose how they contribute and what they learn, and (b) participants' are given choice in the recipes they prepare and perfect and in the direction of any given activity.

In order to design an environment where participants could learn to use scientific practices and pursue ideas that were personally interesting and motivating for them, we designed two activity sequences. First, KSI participants engage in a semi-structured activity sequence to familiarize and scaffold them in engaging in the scientific practices of asking questions, designing experiments, making observations, measuring, sharing results, and drawing conclusions. Second, participants engage in a flexible exploratory activity sequence where they can take the science they learned during the semi-structured activities to iteratively perfect recipes of their choosing.

The overall goal of the semi-structured activity sequence is to engage participants in conducting cooking and science experiments that are focused on understanding what makes foods rise or thicken. Learners and facilitators design experiments that highlight the effects of varying different amounts or types of ingredients in a recipe. The cooking experiments highlight the effects of ingredients (e.g., increased height, volume, density or thickness) in the context of the recipe they are preparing. During these activities, the facilitator helps participants develop a curiosity about the roles particular ingredients (e.g., leaveners and thickeners) play in their recipes by modeling how to use observations to ask questions and to design experiments to answer the questions raised. For example, participants make pudding with different types of starch thickeners and the facilitator helps them to make observations and compare the different textures and thicknesses of the thickeners. These experiments are designed as a whole group. In small groups, learners carry out each variation and then share their results, taste their dishes, and draw conclusions as a whole group.

Semi-structured activities are then followed by flexible exploratory activities called Choice Days where learners are given the opportunity to use what they have learned about particular ingredients to prepare recipes of their choice and make them come out with their preferred taste, texture, and mouthfeel. During these activities, learners have the opportunity to ask new questions, to practice using results and conclusions drawn from the semi-structured experiments to design experiments that answer their questions.

To help bridge learners' goals and the goals of KSI, we position learners first as chefs who want to understand how to make their recipes come out with the desired taste, texture, and mouthfeel. We then transition them into being investigators, who are curious about the roles that each ingredient plays in making their recipes come out as desired. Last, we help them use scientific practices to find answers to their questions about particular ingredients so that they can make informed decisions about the ingredients they use in preparing their dishes. To help them make these transitions, we encourage learners to play with their food as a means of developing more questions, observing their results, making comparisons across variations, and building understanding of scientific phenomena through engaging in scientific practice.

## Methods

### Data Collection

The cases presented in this paper were taken from data collected in our most recent enactment of the KSI program. The program was implemented as a one day a week after-school program as part of a larger weekly after-school initiative by a local YWCA to engage teen girls in science and technology related activities. The participants in the study were from the same suburban middle school where the population was 99% African American. Thus, all participants were African American girls in 6th – 8th grade. Participation varied over the 9-month period, but we had 15-20 consistent participants (7-9 6th graders, 7-10 8th graders, one 7th grader). The program was led by a team of 3 facilitators: the authors and the school program coordinator. For each session, we collected and transcribed video recordings of each group. In addition, after each session we recorded post-observation field notes that captured the significant learning events that occurred during the session. As a part of Tamara Clegg's (Tammy) dissertation on learning and identity, we also conducted and transcribed interviews with four focal learners, their parents, and their science teachers at multiple points during the implementation and once after the implementation ended.

### Case Selection and Data Analysis

The analysis for this paper emerged from a trend Tammy discovered while analyzing the data for her dissertation in the context of exploring the participation of individuals and roles they took on in the program. In particular, she noticed that the roles individuals took on were impacted by the impromptu side investigations participants engaged in and the physical observations they made during these investigation. After noticing this trend, she intentionally identified all the side investigations and physical observations made by small groups of individuals in her data set. She then selected five particularly salient instances of these practices where scientific practice was enhanced and where learners showed or reported deriving personal meaning from the experience. To establish validity, she then presented these instances to the second author, who reviewed the transcripts and videos to ensure they were representative of scientific practices and personal meaning for learners. From these, three cases were selected based on the range of scientific practices participants engaged in and the strength of evidence from transcripts and interviews that suggested the practices were useful and meaningful to the participants. With the unit of analysis being the small group, we then coded each case for instances when groups engaged in the following scientific practices: asking questions, making observations, making explanations of scientific phenomena, sharing results with others, and drawing conclusions.

## Case Studies

As you read the three cases that follow, notice that Cases 1 and 2 represent the ways in which participants initially engaged in simple scientific practices in KSI on Day 12, while Case 3 characterizes the ways in which they engaged in more complex scientific practices on Day 20, the last day of the thickeners unit. Also notice how the facilitation and artifacts in the environment promoted participants' engagement in the scientific practices modeled.

Cases 1 and 2 occurred on Day 12, two days into the "thickeners" unit. On this day, the main activity was a pudding experiment where each group made pudding with a different type of starch thickener to learn about the variety of textures each thickener created. At the beginning of this session, Tammy, a facilitator, provided an overall framing of the types of scientific practices that the participants would be engaging in: making observations, taking measurements, conducting systematic and precise experiments, and sharing results. Following this framing, she introduced the Food Tasting Activity, an activity we designed to help the learners make objective, descriptive observations about food. This activity was meant to prepare participants to notice the differences in texture and thickness of the puddings they were going to make in the subsequent activity.

### Case 1: Describing, Comparing, and Testing Different Store-bought Puddings

The initial big group activity for Day 12 was the Food Tasting Activity. For this activity, we brought in 15 store-bought desserts and pastries with pudding-based fillings for participants to taste, describe, and compare. We felt this activity was needed to help participants develop imagination about the kinds of complex dishes they wanted to prepare (dishes that combine different ingredients and textures) and to help them develop objective ways of describing and comparing dishes. Our goal here was to move learners from simply sharing their opinions about the tastes and textures of food (e.g., it tastes nasty) to making descriptive observations of particular properties of the taste, texture, and mouth feel of the foods (e.g., it is smooth and thick).

Unfortunately the simple act of tasting the foods was not enough of a catalyst to get participants to make descriptive observations of the food they tasted, even when prompted. For one group of participants, however, a major turning point in the activity occurred when Treeva, bored with tasting the food, and probably satiated, said, "I'm sick of eating this right now." In response, Christina, the facilitator, suggested that they focus on testing and comparing the foods instead of eating them. The girls then began looking at different foods on the table and commenting about their thickness and describing in more detail what they liked and didn't like about the foods they were tasting. After tasting the tapioca, Mercedes said she didn't like the way it felt in her mouth. This prompted Candyce to say that she didn't like the flavor because it was sort of lumpy. Mercedes then agreed that she didn't like the feel of the tapioca balls either.

Not yet satisfied with the level at which the girls were comparing the thickness of the foods, Christina, the facilitator, shifted the activity to helping the girls find ways to describe the texture. She suggested developing a test for thickness. In response, Mercedes, who was dipping her spoon in and out of a cup of pudding and watching how it fell off the spoon, suggested testing the thickness by using a spoon to measure how long a particular food stayed on the spoon when turned over. The test sparked Candyce's interest, and she began suggesting foods to try in this test. The two girls then began playfully placing samples of the puddings and fillings on spoons and turning them over at the same time to see which one fell first. Christina then prompted the girls to use this test to rank the food on the table by thickness. She helped them interpret the meaning of their test with respect to the speed in which it fell and which fell first. The girls concluded that the pudding that fell from the spoon first was thinner than the one that fell last. Observing their tests, the eighth grade participants on the other side of the table began to do the spoon test as well. Intermixed in their testing, Candyce and Mercedes started playful competitions based on their predictions about which foods would stay on the spoon longest and which would be the thickest in the ranking. Included in their predictions were descriptions about how the foods looked and smelled. While they ran the test, Christina helped the girls use the results to rank the foods by thickness by rearranging the food artifacts on the table. Once they completed their rankings, the girls enthusiastically demonstrated their spoon test and shared their results with the rest of the group.

Candyce later remembered and recalled her experience with the store-bought fruit tart during the food tasting activity as not particularly pleasant because of its color. In recalling this experience, Candyce without prompting used a descriptive observation about its color to explain why she did not prefer to make her own fruit tart. During later sessions, these learners also became more consistent in making descriptive observations. For example, on Day 20, as you will see in Case #3, Candyce's described the buttermilk as "creamier and thick," and this descriptive and comparative observation led her to wonder about the difference between buttermilk and milk, which, in turn, initiated a buttermilk vs. milk experiment to figure out the answer.

## Case #1 Discussion

*What made this episode scientifically meaningful to the participants?* Candyce's recollection of the fruit tart indicates that she used her observations about the foods to reflect on the types of foods she would be interested in eating again or preparing herself. We also know that this case was scientifically significant for Candyce because we can see that she learned from it. At the beginning of the Food Tasting Activity on Day 12, she, like the other participants, was having difficulty making descriptive observations and comparisons because they were simply focusing on tasting and eating the foods. However, by the end of the episode the girls were making more descriptive observations as a result of engaging with the food as an object with properties that could be tested and ranked.

*How did the enactment promote this?* First, the facilitator allowed participants to engage in the activity as it felt most natural to them, i.e., through playing with and tasting the foods. Then the facilitator recognized a shift in the participants' interest and participation, from interested to bored, and she used this as an opportunity to refocus the participants' attention onto the scientific practices of testing, making observations, and drawing conclusions. Specifically, the facilitator transformed their play practice of dropping the pudding-based foods from spoons suspended in the air into a more formal test for ranking the thickness of the foods they tasted. The facilitator's ability to create an authentic opportunity for engaging the participants in these scientific practices was also created by the availability and diversity of artifacts, namely the pudding-based foods with varying degrees of thickness for which the crude spoon test would produce noticeable differences in the time it took the pudding to fall. In addition, the facilitator played a major part in structuring the activity by keeping track of results, and using the artifacts themselves to represent the rankings the participants were developing and making the rankings available for participant inspection. This representation of the rankings (the puddings lined up according to how thick they were) allowed the participants to confidently share their results and procedures with other KSI participants.

## Case 2: Viscometer Test

Following the Food Tasting activity, the learners divided into small groups to do a simple cooking experiment. Each group made a simple vanilla pudding recipe but used a different type of starch thickener. The over arching goal of the activity was to understand why foods have a range of textures and thicknesses and the roles starch thickeners play in creating those difference. Thus, we wanted the learners to make objective (descriptive and quantitative) observations of the differences in thickness of puddings with different thickeners, so that they could use what they learned to make more complex dishes similar to the ones that they tried earlier during the Food Tasting activity. To achieve this, we planned to have learners measure their pudding with a tool called a *viscometer* that measures the viscosity or thickness of liquids. Using a viscometer, the girls would be able to measure, in seconds, how long it took for the pudding to flow through the opening in the bottom of the viscometer. While the spoon test they had developed during the food tasting activity allowed learners to rank thicknesses and to characterize the thickness qualitatively; the viscometer allowed them to measure thickness quantitatively.

Measuring the thickness of their pudding was difficult because the girls did not finish making their pudding until near the end of the session. They had to wait for it to cool before putting it through the viscometer, and they were running out of time. However, Tammy saw that some of the group members were not as active as others on this team during the pudding preparation. She therefore encouraged those who were not busy with the pudding to measure the viscosity of milk while their pudding was being prepared. Candyce and Mikayla responded to Tammy's prompting and Tammy showed the two learners how to set up and use the viscometer. When the facilitator asked what they thought would happen, Candyce predicted that the milk would come out more slowly than water because milk is "sort of thick," as compared to water.

As they were waiting for their pudding to cool, Amber asked to take a parfait home. Amber's interest in the pudding made Tammy realize that although they could not measure the viscosity of their pudding right then, they could measure and compare the viscosity of other foods they had tasted and examined earlier. Tammy suggested that they try measuring the parfait with the viscometer to see how its viscosity compared to their pudding. When Tammy brought the parfait for measurement, Candyc, promptly took the parfait and began setting up the measurement with Amber. As they measured, Amber excitedly called Tammy over several times because the parfait was taking much longer to come out of the cup than she expected.

Since the girls were able to stay late, they were able to measure the viscosity of their pudding once it cooled. As they set up the viscometer to measure the pudding, Amber and Candyce made predictions about how long it would take. Amber thought their pudding was thicker than the parfait, so it would take the pudding longer to flow from the hole in the viscometer. As they began the measuring the pudding, they were surprised to see that the pudding was not coming out of the viscometer at all. One group member joked that they could walk up and down the hall for an hour and the pudding would not go anywhere. Mikayla excitedly suggested that they try it. Amber suggested they walk up and down the hall three times. The facilitator permitted Amber, Mikayla, and Precious to carry out their test while Candyce held the viscometer. Upon returning, Amber immediately checked the viscometer and said, "Okay, it's still not done, is it?" When she saw that no pudding had flowed from the viscometer, she exclaimed, "Knew it!"

Later, we learned from Amber that she told her science class about her experience measuring viscosity in KSI when they discussed the topic of viscosity in class. She also talked about the viscometer test, three months later, on Day 20, when the idea of viscosity came up in presentations to their parents. Later, during an interview, she told us that the viscometer measurement was something that she did as a scientist in KSI.

### Case #2 Discussion

*What made this episode scientifically meaningful to the participants?* Participants were able to move from descriptive observations to quantitative representations of thickness when encouraged by the facilitator. In this way, learners were moving from simpler forms of scientific practice (i.e., making observations and predictions) to a more sophisticated form (i.e., using observations to support predictions, initiating their own tests, and designing creative ways to explore their predictions). Amber seemed to be proud of these accomplishments and to see them as scientific, as evidenced by the three different times we recorded her sharing the viscometer experience and its significance to her with others.

*How did the enactment promote this?* Initially, the facilitator recognized the opportunity to engage learners in the scientific practice of taking quantitative measurements to compare thicknesses of the pudding, yogurt, and milk. She saw these opportunities when the participants (1) were bored and (2) were interested in tasting a new dish. Then she introduced learners to quantitatively measuring thickness by showing them how to use the viscometer to take a scientific measurement of thickness. Finally, the facilitator's recognition of the availability of the viscometer, milk, and parfait prompted their impromptu comparison of the thicknesses. It is also important to notice that the parfait was available and accessible from a previous activity (the beginning food tasting discussion) even after its intended purpose was past.

## Case 3: Milk vs. Buttermilk Investigation

Day 20 was the final day of the program, in which we invited the learners' parents and family members to come for an open house during the second half of the session. On this day, we wanted participants to use their knowledge of leaveners and thickeners to revise a recipe of their choosing to achieve their personal goals for the dish. Candyce, Treeva, and Rachael decided to make chocolate cake. Prior to cooking, they thought a little about the leaveners and thickeners they would use in their recipe, but it was not until they were baking and wondered about the difference between buttermilk and whole milk that they really began to think about leaveners in their cake.

As Candyce and Rachael measured buttermilk for their cake batter, Candyce, noticed that buttermilk was "creamier" and "thicker" than milk. She asked about the difference between buttermilk and whole milk. Tammy told her how buttermilk is made and that when buttermilk and baking soda are mixed, a foaming reaction happens. Then, when she realized that we had milk, buttermilk, and baking soda available to help her demonstrate the reaction, she began to set up what she was describing. As Tammy measured the baking soda and whole milk, she asked Candyce and Rachael what they thought would happen. Rachael predicted that the mixture would not bubble up. After adding the milk to the baking soda, they saw that nothing happened.

Switching to the next variation, Tammy asked what they thought would happen when they added buttermilk. Rachael thought it would bubble. Tammy reminded the girls about the baking soda and baking powder experiments they did several months before, and she asked them what type of ingredient they needed to mix with the baking soda in order for it to begin producing air. Treeva remembered it needed an acid. When Tammy poured the buttermilk in the glass, they were surprised to see that nothing happened to the mixture at first. Tammy then noticed something at the bottom of the buttermilk glass and they saw the mixture start to bubble up. Amazed at the size and look of the bubbles, all the girls leaned in over the table to see what was happening in the glass. Treeva picked up the buttermilk glass and asked, "What is it doing?" Candyce made the observation that there were bubbles at the top and that it was foaming. Based on their curiosity about the reaction, Tammy explained that the mixture was producing air, meaning $CO_2$, because buttermilk is an acid. Treeva, became excited by the smell of the mixture; she and Candyce agreed it smelled like baby milk.

As the group continued making the cake, they periodically noticed changes in the smell and look of the buttermilk glass, playfully making observations about it. At one point, Candyce noticed the glass sitting on the table and excitedly said that it looked strange, noticing that the baking soda seemed to settle on the top of the buttermilk. She dipped her fingers into the cup and walked over to Treeva (who was mixing the cake batter), showing her the buttermilk mixture on her finger. Later, when Candyce mentioned throwing the mixture away, because it looked "creepy," Tammy told her she could save it and show their parents in their presentations.

During presentations to their parents, when the facilitator mentioned baking soda and baking powder, Candyce immediately brought up their experiment they had done and walked across the room to get their buttermilk mixture to show the parents. She and Rachel reminded facilitators of the experiment during the presentations when they were about to move on to the next group without allowing them to discuss it. With prompting from Tammy, Candyce explained the concept of leaveners, and Rachael and Candyce described their experiment and results. Afterwards, Candyce moved around the room showing each audience member and fellow KSI members the buttermilk mixture. Later, in a subsequent interview, Candyce used the buttermilk and baking soda experiment as an example of how science has helped her figure out why things happen in recipes:

> Candyce: okay if you have some ingredients, and let's say you mix them together, but something happens and you don't know what it is, then that's where science can help you. You can figure out like why this happened. Like say you mix, baking soda and buttermilk together. You could say like this happened because buttermilk is an acid.

### Case #3 Discussion

*What is scientifically meaningful about this enacted activity?* Learners engaged in the scientific practices of asking questions, making observations, and monitoring results. These experiences also contributed to their understanding of leaveners (as evidenced by their later explanations). We also see that this experiment was meaningful for learners in their continued interest in the artifact even after seeing the initial reaction. For example, they initiated and sustained discussion of the experiment later with their parents. We also see that reflection on the experiment in subsequent interviews helped Candyce see the connection between cooking and science.

*How did the enactment promote this?* First, the facilitator recognized Candyce's curiosity about buttermilk as an ingredient. Then she recognized this curiosity as an opportunity to provide the appropriate prompts, explanations, and experiments to help the group answer their questions. Next, she modeled the practice of using experiments to make explanations by leading an experiment using available materials to show learners the differences between buttermilk and milk. She also gave learners opportunities to participate in the experiment by prompting them to make predictions and discuss what they already knew about the ingredients. Again, the diverse artifacts we had available for this experiment also enabled the experience. Although their recipe did not call for whole milk, it was still available in the environment and the facilitator was aware of that when she initiated the experiment. Furthermore, the unusual reaction of buttermilk with baking soda enhanced leareners' interest in making observations and later sharing their results with the whole group and their parents.

## Discussion

*What is a scientifically meaningful experience for middle-schoolers?* Earlier, we defined scientifically meaningful experiences as ones in which learners make scientific accomplishments that have meaning relevant to their lives. Over the span of the three cases presented here, we see that learners were able to participate more authentically as scientists over time, in that they began to make predictions and descriptive observations, using descriptive observations as evidence for their predictions. Learners also made quantitative measurements of their results (i.e., their dishes). Finally, they were able to make scientific explanations, sharing their results with others. Chinn & Malhotra (2001) suggest that these practices are authentic to the practice of scientists and should be promoted in school science. Here we illustrate how an informal environment supports middle school learners in engaging in these practices. Relating these accomplishments to learners' lives, we see that learners were able to share their accomplishments with their peers and parents. In subsequent interviews, we see that these accomplishments contributed to learners' reflections on their participation in science and on the relevance of science to cooking. Thus, the scientific accomplishments learners made in the context of their everyday lives helped them to think about their role in science and the relevance of science in their lives.

*When does our design promote the having of meaningful scientific experiences*? We can see from our results that it is hard to predict when learners will begin to ask questions and make connections from explanations. Intended goals were not always realized when we expected them to be (if at all), but many were realized over time. Even as they were, they did not happen in the way we expected, nor the timeframe, as we might expect based on Crowley & Jacob's (2002) work. In KSI, the design of earlier experiments and activities was intended to provide common experiences to refer back to over time to help learners perfect their recipes. Facilitators were able to guide participants back to thinking about these experiences later because all activities were designed to contribute to participants' achieving a larger goal that learners had and because learners were aware of what each activity could contribute to achieving that goal. For these reasons, it was natural for facilitators to bring learners back to those experiences at relevant moments, just as Crowley & Jacobs (2002) advise. But, while Crowley and Jacobs help us understand how to use timely opportunities to provide children with explanations and to help them connect across experiences, our work uses the same ideas to suggest ways of finding timely opportunities for learners to explore answers to their own questions and use those experiences to help them connect their experiences.

*How can we promote the having of those experiences?* This work offers a glimpse into the underlying mechanisms of facilitator support and artifacts in an informal learning environment where facilitators may be less aware of participants' attention, interests, and curiosities than a parent might be with their child while visiting a museum (Crowley, 2002). Yet, in the cases we presented, facilitators were able to recognize times when learners' interests and attentions were piqued and when they were dwindling. They were then able to capitalize on these moments, using them as opportunities to engage learners in unplanned scientific activities by selecting the appropriate scientific practices for the given moment. Keeping track of learners' interests and attentions and the learning opportunities available may seem like an impossible feat. However, when working with small groups of learners, facilitators can often build off of the interest of one participants to help others become interested and engaged.

To promote such unplanned scientifically meaningful experiences, facilitators must also be equipped with the necessary artifacts to help the learners increase their understanding, familiarity, and experience with the scientific practices (e.g., the puddings with a range of textures). And they must take advantage of the opportunities for piquing interest when the opportunities arise rather than waiting for lulls in activity. Our work shows that, unlike museum visits or home explanations, artifacts do not only trigger the explanations; they can also be used to engage learners in building, interpreting, and applying explanations.

## Conclusions

We presented an approach to helping participants learn and practice science in the context of their own interests and goals. The approach led to scientifically meaningful to experiences for learners. We found that helping learners to play with artifacts promotes learners' engagement in authentic scientific practices and meaningful experiences. But these experiences must be grounded in the context of a larger goal that learners have, and facilitation is needed to orient learners' practice with the objects, helping them to explore and investigate in meaningful ways. More work is needed to understand the complexity of these learning environments and to understand the impact these meaningful experiences have on learners' identities and views of science.

## References


Chinn, C. A., & Malhotra, B. A. (2001). Epistemologically Authentic Scientific Reasoning. In K. Crowley, C. D. Schuunn & T. Okada (Eds.), *Designing For Science: Implications From Everyday Classroom, and Professional Settings* (pp. 351-392). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Clegg, T., Gardner., Williams, O., & Kolodner, J. (2006). Promoting Learning in Informal Learning Environments. S. Barab, K. Hay, & D. Hickey (Eds.), Proceedings of the Seventh International Conference of the Learning Sciences (pp. 92-98), Bloomington, IN.

Clegg, T. L., & Kolodner, J. L. (2007). Bricoleurs and Planners Engaging in Scientific Reasoning: A Tale of Two Groups in One Learning Community. Learning Communities Special Issue of Research and Practice in Technology Enhanced Learning, 2(3), 239-265.

Crowley, K., & Jacobs, M. (2002). Building Islands of Expertise in Everyday Family Activity. In G. Leinhardt, K. Crowley & K. Knutson (Eds.), *Learning Conversations in Museums*. Mahway, NJ: Lawrence Erlbaum Associates.

Gleason, M. E., & Schauble, L. (1999a). Parents' Assistance of Their Children's Scientific Reasoning. *Cognition and Instruction, 17*(4).

Gleason, M. E., & Schauble, L. (1999b). Parents' Assistance of Their Children's Scientific Reasoning. *Cognition and Instruction, 17*(4), 343 - 378.

Nasir, N. (2002). Identity, Goals, and Learning: Mathematics in Cultural Practice. *Mathematical Thinking and Learning, 4*(2&3), 213-247.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2001). *What 'Ideas-About-Science' Should be Taught in School Science? A Delphi Study of the Expert Community*. Paper presented at the Sixth International History, Philosophy and Science Teaching Conference.

Resnick, M., Bruckman, A., & Martin, F. (1996). Pianos not stereos: Creating computational construction kits. *interactions*, 3(5), 40-50.

Rutherford, F. J., & Ahlgren, A. (1991). *Science for All Americans*. New York: Oxford University Press.

Seiler, G. (2001). Reversing the "standard" direction: Science emerging from the lives of African American students. *Journal of Research in Science Teaching, 38*(9), 1000-1014.

Shaffer, D. W., & Gee, J. P. (2007). Epistemic games as education for innovation. *British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education - Current Trends, 1*(1), 71-82.

Shaffer, D. W., & Resnick, M. (1999). "Thick" Authenticity: New Media and Authentic Learning. *Journal of Interactive Learning Research*, 10(2), 195-215.

# Science Learning as the Objectification of Discourse

**Abstract:** The study presented here seeks to contribute to the dialog of how students come to learn disciplinary knowledge. Students in a highly-interactive physics course were studied as they constructed a model of magnetism. Two focus students were video taped throughout the 7-hour unit. They transitioned through 7 different representations of their models and their discourse and model-building practices evolved. The findings of this study are framed in terms of Sfard and Laive's (2005) ideas about the objectification of discourse as a definition of learning. The findings from this study are used to demonstrate the power of this perspective and to illustrate that the learning of disciplinary knowledge resulted from the students objectifying their discourse away from the self and its communication with other people and toward the self in relation to the human-independent world of magnetic phenomena.

## Background and Theory

Language is often treated as a communication medium rather than as a legitimate part of cognition in research on the learning of disciplinary knowledge. Many researchers have studied the role of language in the development of scientific knowledge (Lemke, 1990; Moje, Collazo, Carrillo, & Marx, 2001; Kelly, Crawford, & Green, 2001; Lee & Fradd, 1998) and in other types of disciplinary knowledge (Gutiérrez, Lopéz, Alverez, & Chiu, 1999; Gutiérrez, Lopéz & Turner, 1997; Gee, Allen, & Clinton, 2001; Gee, 1990). It is ever clear in this body of research that both language and concepts are involved in the learning process but they are very difficult (if not impossible) to analytically distinguish from one another. Roth and Duit (2003) refer to this phenomenon as the "structural coupling" of language and concepts, a phenomenon that remains a largely unexamined field of inquiry and is the subject of the research reported in this paper.

Gee uses the term "Discourse" not to refer to the exchange of words in general, but to refer to the language and practices of a particular community. In science, Discourse includes the formal terms and how terms are put together in sentences (e.g. an object does not *have* a force, an object *exerts* a force), symbols and their uses (e.g. an arrow on top of an the letter *a* means something different from arrows extending from a dot), what is deemed salient in a given situation (e.g. differentiating "noise" and signal on a graphical representation), the practice of making evidence-based claims, mechanistic reasoning, and the practice of creating and using models of phenomena. The preceding list of scientific practices constitutes much of what is meant by *scientific literacy,* or being literate in science. Like any literacy, it consists of a set of obligations, expectations, practices, values, meanings, and ways of using language that make up participation within the community thus defined. Within the scientific community this includes the learning of scientific terms and how to use them, the symbols and their contextual applications, how to read out the appropriate features of a given situation, and how to argue mechanistically in support or rejection of a model of an observable phenomenon.

Although models of phenomena are central to all scientific discourse, research on students' understanding of models in science has only begun to blossom in the recent decade. Work on meta-modeling competence (Schwarz & White, 2005), students' understanding of the nature of models and modeling (Harrison & Treagust, 2000; Harrison & Treagust, 1998; Gilbert & Butler, 1998; Windschitl, 2004; Windschitl & Thompson, 2006; Snir, Smith, & Raz, 2003; Bhushan & Rosenfeld,1995; Harrison & Treagust,1998; and Coll, France & Taylor, 2005), and students' development of mechanistic reasoning strategies (Russ, Scherr, Hammer & Mikesha, 2008) are increasingly becoming part of the mainstream of science education research. Overall, these studies continue to show that by in large, science classes fail to explicitly teach, and students fail to learn skills for developing, revising, and utilizing scientific models to explain observable phenomena. This finding is critical because models and modeling are the crux of the Discourse of science.

The practice of reasoning with models and mechanism involves testing for a hypothesized process involving underlying *mechanisms* that could drive observable phenomena (Machamer, 2004). In a study, Windschitl (2004) demonstrated that science majors revealed a common "folk theory" of scientific inquiry that did not involve model-based or mechanistic reasoning. The secondary science teacher certification students in his study, "did not make the methodological connection that investigations should be based on some explanatory premise nor was there evidence that they understood that the goal of inquiry is to support, revise, or refute various aspects of scientific models" (p. 491). Instead, they engaged in "relation-based reasoning" where empirically testing relationships between variables was viewed as an epistemological end in itself. Windschitl argued that these students viewed models and theories as optional tools that could be used to help explain results after a scientific study is complete. In contrast, much of science *is* the practice of building models and experiments serve the purpose of demonstrating that a model is adequate for explaining phenomena. High school and college students have difficulty developing skills associated with model-based reasoning (Windschitl, 2001; Abd-El-Khalick, 2001) and often fail to learn the Discourse of science.

In a provocative publication, Sfard and Lavie (2005) argue that learning within a discipline (in their case mathematics) involves the *objectification of discourse*—the use of words as if these words signify *discourse-independent entities* out there in the mind-independent world (Sfard & Lavie, 2005). Sfard and Lavie paraphrase Vygotsky and go on to indirectly define learning as a process that "begins as an interpersonal affair (and) turns in the growing mind, into a matter of one's relation with human-independent world." They also claim that, "…this kind of development is a 1-way process, and the change from the interpersonal to between-person-and-the-world outlook, once accomplished, can hardly be reversed," (p. 238-239). This is Vygotsky's theory of concept formation (Vygotsky, 1986), and in line with Vygotsky's theory, Sfard and Lavie (2005) define learning as an irreversible change in outlook that requires the "ability to see as 'the same' things that, so far, could only be seen as different," (p. 238). I will elaborate on this idea throughout this paper, but first a note on the notion of *disciplinary knowledge.*

Human beings tend to talk about disciplinary knowledge as if it is a thing that a person can possess rather than as a narrowly defined set of obligations and expectations that form the basis for communication (synchronous and asynchronous) with individuals who are familiar with, identify with, participate in, and continue to define these obligations and expectations. In the 1930's Ludwik Fleck developed the idea of *Denkkollektiv* (or *thought collectives*) and in 1935 he carefully elaborated this idea in his book, Genesis and Development of a Scientific Fact (Fleck, 1975). Fleck used the term thought collective to describe a system of (ever evolving) knowing that participants draw from and contribute to, a system that exerts a compulsive force on the thinking of an individual and thus constrains the set of possible thoughts of individuals within a community, but not on the community itself. Similarly, Sfard and Lavie (2005) describe disciplinary knowledge in terms of "endorsed narratives," which are defined as "sets of propositions that are accepted and labeled as *true* by the given community," (p. 246). The work of Sfard and Lavie (2005) synthesizes nicely with Fleck's work in attempts of answering the shockingly complex question: *what is disciplinary learning*. Sfard and Lavie argue that because human language is ontologically and epistemologically laden, it has up to now been difficult to understand the learning process due to the fact that disciplinary terminology is often (if not always) a result of the irreversible *objectification* that they claim defines learning. In other words, since language lives in the space of "the learned" it is very difficult to apply to the space of the "unlearned" or the space of the "learning," because each space is replete with its own (somewhat distinct) set of obligations, expectations, and practices that do not necessarily lie along a continuum. In fact, an entire literature on misconceptions in science has been established that merges these two spaces (for a discussion see Otero & Nathan, 2008). Indeed, Vygotsky's theory of concept formation mostly applies to the space of "learning," and throughout his influential work he struggles with language in efforts of bridging the space of the "unlearned" and the space of the "learned." The work of Sfard and Lavie brings the field of the learning sciences closer to bridging this gap.

Sfard and Lavie differentiate between disciplinary words and routines. Disciplinary words such as *force* in physics or *smaller* in mathematics have specialized meanings within the respective Discourse. Routines are ways of approaching tasks and come in the form of *deeds*, *exploration*, and *rituals*. *Deeds* are actions actually taken by the individual that produce a change in the environment. For example, if a teacher asks a child to pick the box that is biggest, the child reaches out and moves the larger box closer. *Explorations* have the aim understanding the world. Explorations often consist of extradiscursive entities that are the subject of discourse (as is the case in reasoning with mechanism about an invisible entity such as a charge). Exploration is what science education researchers would refer to as model-based, mechanistic reasoning. Rituals on the other hand, are socially oriented, with the aim of fostering solidarity with whom they are performed. Rituals have the sole aim of communication with the other. These distinctions become important as we attempt to understand science learning through the lens of the objectification of discourse. In observing students as they learn the practice of model-building and reasoning with mechanism, we are looking for transitions from discourse that is solely ritual to discourse that involves exploration and discussion about extradisursive entities.

In the remainder of this paper, I use empirical data from a physics course for adult learners to illustrate how the perspective articulated by Vygotsky (1986) and extended by Sfard and Lavie (2005) can begin to answer the question, "what is science learning?" My perspective differs from that of Sfard and Lavie only in that when they speak of the objectification of discourse, they apply it only to learning that differentiates children from grown-ups. I use this notion more broadly, to apply to all humans engaged in the process of learning disciplinary knowledge. In the research that follows, the modest question, "what does learning look like?" is mapped out through the exploration of three specific questions: (1) How do students' representations of models of magnetism evolve over the unit? (2) How do students' model-building practices evolve over the unit? (3) How do students' language practices evolve over the unit on magnetism?

## Study Context and Participants

Physics and Everyday Thinking (PET) is an introductory college level inquiry-based physics course designed to meet the needs of adult learners, non-physics majors, especially elementary and middle school teachers. This study focuses on the Models of Magnetism unit, throughout which students have several opportunities to develop and revise explanatory models of magnetism with the expectation that through guided experimentation, negotiation, and model revision, they will arrive at an expert-like model of magnetism.

Each activity in the curriculum consists of three parts: initial ideas, collecting and interpreting evidence, and summarizing questions. In the initial ideas section, students are asked to express their prior ideas and thoughts about the phenomenon that will be explored in a particular activity. Throughout the *collecting and interpreting evidence* parts of activities, students do experiments such as rubbing iron nails with magnets and investigating phenomena with laboratory apparatus and computer simulations. Questions in the curriculum frequently ask students to explain how their conceptual models account for their observations or how they might need to revise their models on the basis of observations. In some parts of the unit, *summarizing questions* explicitly guide students through the process of mechanistic reasoning, for example, one summarizing question asks, "Based on the results of the experiment, how would you describe each of the individual entities inside the magnet-rubbed nail?" Intentionally or not, this question asks students to begin to objectify the discourse by introducing the notion that there must be something (entities) inside the nail that is worth discussing. The curriculum is built on research on how students learn, so it is not surprising that by this time in the curriculum, most students have already introduced some type of entity within their drawings. However, the former question and several that follow it explicitly direct students to consider the term "entities" as a pointer to *discourse-independent entities* out there in the mind-independent world. Questions that follow within the curriculum begin to help students to use these entities as the agents that ultimately are responsible for the outcome or observation. Whether the curriculum developers intended it or not, the questioning sequence helps students to objectify their discourse.

A classroom that uses the PET curriculum is an appropriate context for a study of this type because it relies on students' making sense of scientific phenomena by constructing models that can be supported by evidence. There is no textbook used for the course, students are expected to construct their own understandings through small group work and discussion and larger whole-class discussions. The role of the teacher is mostly to lead discussions, making sure that students support claims with evidence, consider all of the evidence when making a claim, and ultimately come to consensus as a class on the ideas that will be used to explain the data.

## Study Design and Method of Data Analysis

This study focuses on a group of two students who were representative of the class, Brie (an experienced elementary teacher) and Mona (a pre-service elementary teacher), who were video taped throughout the course. Both Brie and Mona are white, middle class adults, both with families of their own and neither identified herself as a "science person" at the beginning of the course. The Models of Magnetism unit was approximately 7 hours, all of which was videotaped, transcribed and analyzed. A constant comparative method was used to generate and revise initial codes to characterize the nature of all of the discourse. On the basis of analysis and lengthy discussion by three different researchers, the following codes emerged: metacognitive discourse, explicit discussions about the nature of science and the nature of models, analogical reasoning, evidence-based reasoning, experience-based reasoning, model-based reasoning, conducting experiments, logistics, and off task discussions. All of the data were coded independently by the three researchers who met weekly over the course of one year to compare the coding of each transcript. Only those codes for which a full consensus could be established were kept and analyzed further. Of the consensus codes, all but the last three, "conducting experiments, logistics, and off task" were considered "sense-making" behaviors and were further investigated in order to understand how students' learned the disciplinary knowledge associated with a model of magnetism. Sense-making data constituted 165 minutes, approximately 40% of class time. Many episodes of what we referred to as "finding terms" were identified by all researchers, where the students struggled with terminology and seemed to be looking for the appropriate scientific terms to express their ideas. We therefore re-examined all of the sense-making data for the frequency of use of scientific terms such as force, charge, positive, and energy. The frequency measurements, the characteristics of the discourse, and the models of magnetism were then compared for the purpose of understanding how the students revised their models and how the nature of their scientific discourse changed in the process. Finally, these data were analyzed in terms of the routines used by the students, whether they were deeds, explorations, or rituals. Inferences about objectification of discourse were then made, and supported by changes in the nature of the students' talk as they moved closer and closer to the type of model associated with the endorsed narrative of magnetism in science.

## Findings and Analysis

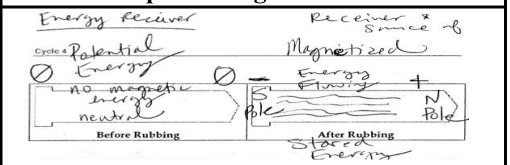
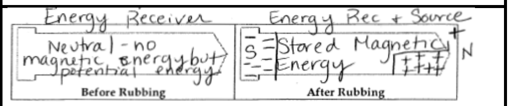
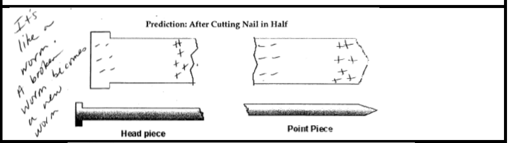
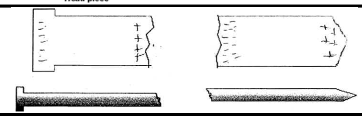
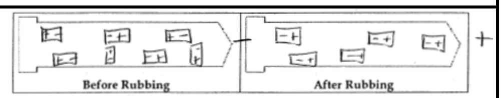
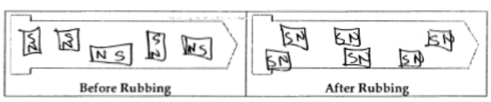
A summary of the findings of this study is organized in four sections. The first section presents the students' representations as they changed over the unit. The next three sections are organized according to three theory-laden claims that outline how discourse was objectified throughout the student's learning process and illustrate that disciplinary learning begins "as an interpersonal affair (and) turns in the growing mind, into a matter of one's relation with human-independent world."

## Evolution of Students' Model Representations

The six models of magnetism that Brie and Mona wrote, discussed, and shared are described in Table 1 in the order that they occurred. Model is defined here as a representation constructed by the students along with the actual statements that they made when drawing or discussing the representation. The drawings in figure 1 were photocopied from the written work of the students. The text used to describe the group's models including the model names, were developed by the researchers but checked with the students after the course was completed. The models described below are not intended to serve as a one-to-one mapping of what was going on inside the students' heads. Instead, they illustrate the models that a group of students co-constructed and presented to each other and to the class.

In the three units that preceded the magnetism unit, the students investigated forces, different forms of energy, and energy transfers. The Energy model was Brie and Mona's initial model of magnetism (model 1), containing much of the same language that was used in the previous units of the curriculum. By the end of the magnetism unit, Brie and Mona developed a tiny-magnet (domain-like) model (model 7), which was the target model of the curriculum and aligned with the scientifically accepted (domain) model of magnetism.

Table 1: Brie and Mona's Models of Magnetism in the order that they occurred throughout the unit

| Model Description | Actual Group Drawing |
| --- | --- |
| **Model 1: Energy Model.** An unmagnetized nail has no energy but it has the potential to have energy, and a magnetized nail has stored energy (usable energy), which becomes magnetic energy during interactions. |   |
| **Model 2: Energy/Charge Separation Hybrid.** An adaptation to the Energy Model, where collections of positive and negative charges now appear at either end of the rubbed nail. |   |
| **Model 3a:** Worm Regeneration Model. A magnet is like a worm insomuch as its properties are replicated on both pieces when it is cut in half. The properties that are replicated are that one end is positive and the other is negative. |   |
| **Model 3b: Charge Separation Model**. Positive and negative charges exist randomly within unmagnetized nails. During rubbing, the charges become separated at the two ends of the nail. |  |
| **Model 4: Fractal Model:** Entities exist in the nail and they are tiny versions of the nail, like an iron filing is to a large piece of iron (nail). The nail takes on the properties of the tiny iron filings. | *The students provided no illustration of this idea.* |
| **Model 5: Activation Model**. Entities exist in the unrubbed nail and are activated when the nail is rubbed with a magnet. The activated entities are represented the same way a magnetized nail is represented, by a plus or minus symbol on either end. |   |
| **Model 6: Plus/Minus Entity Alignment Model**. The positive and negative charges within a magnet are coupled and inseparable. They are randomly arranged in an unmagnetized nail and by rubbing the nail with a magnet, the entities are aligned. |   |
| **Model 7: North/South Entity Alignment Model.** This is similar to the plus/minus alignment model except the entities in the nail are tiny magnets. This is the target model of the unit and consistent with the domain model. |   |

The nuanced discussion of each of the seven models shown in Table 1 is too lengthy to include here. Instead, I present a discussion of the claims that can be made from these models and the associated discourse.

## Claim 1: Models 1, 2, and 3a are illustrative of ritual routines

Brie and Mona's early models *describe* their observations. The energy model (model 1), the energy-charge hybrid model (model 2), and the worm regeneration model (model 3a) all ascribe *conditions* that can *account* for the observed phenomena (attraction, repulsion, and the dipole behavior of both ends of a cut nail). None of these three models provide a mechanism for how these macroscopic observations could come to be nor do they provide discussion about the properties of the entities (the little plus and minus symbols that appear in model 2 and 3a) beyond that they represent opposite things.

I argue that, for Brie and Mona, the purpose of the first three models was to mediate a discursive procedure (ritual routine) rather than to understand the world (exploration routine). This is not to say that Brie and Mona's discourse does not represent their thinking about the phenomenon. On the contrary, they clearly have ideas about energy and are trying to describe the behavior of the magnetized and unmagnetized nail. The transcript and the group's representations suggest that the visual mediator of the discourse around models 1, 2, and 3a is the nail itself. The pronoun "it" is used by the students to refer to the nail (later "it" will be used to refer to the entities within the nail.) Also, a majority of the discourse was an attempt to determine the appropriate words to use. This was partially constrained by the situation; the students were expected to express their ideas first in their workbooks and then on a 3'X4' dry erase board and present it to the class. But in both cases it seems that they were preparing their ideas for others (the teacher or the other rest of the class) to describe *that* a nail becomes magnetized rather than seeking to understand *how* a nail becomes magnetized. The excerpt below demonstrates that Brie and Mona searched for the correct terminology. Some transcript has been removed for brevity, represented by ellipses.

| | | |
|---|---|---|
| 500 | M | What do we:: agree on? Yeah negative and positive. So we agree that |
| 501 | B | Something's flowing in it |
| 502 | M | And that it's become magnetized. Or, has magnetic energy, right? |
| … | | |
| 506 | M | Um:: I like the idea that it has two opposing ends, or forces, or whatever you want to call it |
| 507 | B | So south and north? |
| … | | |
| 508 | M | Yeah, or just somehow maybe indicating that this one is neutral, it doesn't have, you know what I mean? Does that seem right? |
| 509 | B | Yeah, what is neutral? What's the sign for neutral? S? |
| 510 | M | Mmm, or a positive N? |
| … | | |
| 518 | B | Well, yeah. Neutral means nothing. And that's like, don't do anything, |
| 519 | M | No magnetic energy |
| 520 | B | Does not do anything. You know? |
| 521 | M | Doesn't have any energy? |
| 522 | B | ((makes a circle with her hands)) No smoking |
| 523 | M | Has energy potential but it doesn't have any energy. |
| 524 | B | Let me write that on my thing, I kinda like that |
| … | | |
| 536 | B | You're right. You said a word, now what is it, what was it? |
| 537 | M | I just said that energy was present, but energy is *infused*, I don't know. |
| 538 | B | Potent... oh no, not potential. |
| 539 | M | Stored! |
| 540 | B | ((waves her arms and legs around excitedly)) |

In lines 500-507 the students are looking for a way to represent that the nail that has been rubbed with a magnet is magnetized and the unrubbed nail is not. They use terms such as negative and positive, magnetic energy, south and north, and neutral. In lines 518-522 they search for the term that would be necessary to describe the concept of *nothing* and conclude using the terms *neutral, no magnetic energy, potential energy,* and a circle with a line through it. In lines 536-540 the purpose of the discourse appears to be to negotiate which words would most appropriately describe an agreed upon condition of magnetized versus unmagnetized. These terms are also present on their representation of Model 1 in Table 1. Brie and Mona revised their model to include some of the symbols used by the other students (pluses and minuses) for model 2 and developed a worm regeneration analogy for model 3a when they were asked what would happen when they cut the magnetized nail in half. In all three models, Brie and Mona discussed only the nail in a broad sense and not what was happening within the nail. I argue that this represents a ritual routine because the discussion centers on how to *describe* the macroscopic properties of the observation—while privileging scientific words, rather than on *explaining* the observations themselves. The rest of the models were, however, mechanistic.

## Claim 2: Models 3b, 4, 5, 6, &7 illustrate the exploration routine

I argue that models 3b-7 represent exploration routines (model-based, mechanistic reasoning), although the students were still learning how to engage in exploration as they were doing it. Brie and Mona's discussion about each of these 5 models had the visual mediator of the entities inside the nail, rather than the nail itself. The words "it" and "they" no longer referred to the nail; they now referred to the entities inside the nail. These entities were "extradiscursive" in the sense that they were treated as if they existed independent of the discussion or the classroom activity, in the human-independent world. These entities were not intended for the purpose of communication. In these cases, the purpose of the discourse was to negotiate an understanding of *why* the nail behaved the way it did. The discourse was used to discuss the properties, activities, and organization of the entities and ultimately how these things impacted the behavior of the observable nail. As is evident in Table 1, all of the energy terms have vanished by model 3b, and instead plus and minus symbols are used throughout the remainder of the discussion until model 7, where Brie and Mona explicitly switched to "Ns" and "Ss" due to an activity that drew a distinction between electrostatic and magnetic phenomena. Brie and Mona were aware that they did not know what plus and minus really meant except that they were opposites and therefore could be used to signify attractive and repulsive behavior. When the students decided to switch from singular plus and minus entities (models 3b) to dipole entities (models 5-7), Brie and Mona decide to try out the term "electron" to stand in for the objectified entity that was the subject of their discourse. The point is that in later discourse (presented below), Brie and Mona were no longer negotiating which term to use, they were negotiating the behavior of the thing that they used the term to describe.

| | | |
|---|---|---|
| 2268 | B | Let's call them electrons just for kicks and see if we're right. |
| 2269 | M | Ok. |
| 2270 | B | So each electron in the nail behaves the same |
| 2271 | M | I don't know my electron. I wish I had a defin... I should have looked up electron, you know to define it. I wonder what that word means. |
| … | | |
| 2279 | M | Ok, so |
| 2280 | B | It takes on the same... |
| 2281 | M | Right, same properties. Right. |
| 2282 | B | The nail |
| 2283 | M | The nail |
| 2284 | B | ((writes)) takes on the same properties |
| 2285 | B | and...there's something else, and behaves |
| 2286 | M | and the same, um |
| 2287 | B | and behaves in the same way, 'cause don't they behave in the same exact way? |
| 2288 | M | mmhmm yeah. Each one of those behaves in the same way as the nails does, and the same way as a piece of the nail did. |
| … | | |
| 2293 | B | because it takes on the properties and it behaves the same way, you have to say both. |

The discourse above took place as the students were developing model 4, the Fractal Model. As soon as they began to discuss the actual behavior and properties of the entities they moved back to plus and minus symbols and terminology. The difference between the discourse in lines 500-540 and that in 2268-2293 is that the purpose of the former was to put words onto already agreed upon experiences—to communicate with one another, the teacher, and the other students—that the rubbed nail was magnetized and the unrubbed nail was not. In the latter (2268-2293), the purpose was to negotiate the properties and behaviors of invisible entities that might exist within the nail and help to explain the nail's behavior. Later transcript reveals how Brie and Mona chain back and forth between the microscopic entities and the macroscopic behavior of the nail.

## Claim 3: The curriculum shaped the students' discourse and hence, their reasoning

The data suggest that transitions between Brie and Mona's models were facilitated by specific constraints provided by the curricular materials both in the language that was used in the workbook and in the experiments that were prescribed. An experiment that led to the generation of Model 4, provided an analogy to the nail. A closed test-tube was about ¾ filled with iron filings and students rubbed it with the magnet just as they did the nail. They observed similar macroscopic effects. They could also see the effect of rubbing the test tube (analogical nail) on the little iron filings inside the nail. This helped to shift students' attention from the nail as a macroscopic object to the entities within the nail. Further, a summarizing question asked, "Based on the results from the experiment magnetizing the test tube filled with iron filings, how would you describe each of the individual entities inside the nail." Although Brie and Mona had already been talking about entities in the form if plus and minus charges for a while, they now began to focus on the properties and behaviors of these entities. In doing so, the language that they used to describe their model shifted from discussing the magnetized nail as an "it" that behaved differently as a result of being rubbed by the magnet, to discussing the magnetized nail in terms of "they," the little things inside the nail and what they were doing as a result of the nail being rubbed by the magnet. This shift in attention refocused the discourse from the aim of communication to the aim of understanding the behavior of the entities. This

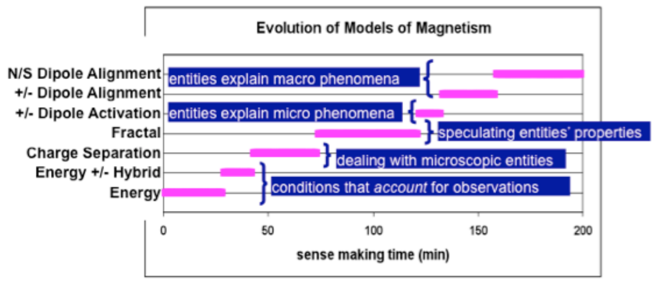



Figure 1: Students models (along y-axis) over time (x-axis) and their mechanistic reasoning.

led to discussions about the how the behavior of the entities could account for their macroscopic observations, which often led them back to change the behavior and properties of the entities until the Brie and Mona arrived at a fully mechanistic model that could account for the observed phenomenon. This backward and forward chaining between mechanistic model and observable phenomenon has been observed by other researchers who study mechanistic reasoning (Russ et al., 2008). Figure 1 summarizes how the students' models evolved along with their attention to different aspects of the entities.

Figure 2 provides another representation of the changes in Brie and Mona's discourse patterns throughout the unit. In figure 2, two 13-minute episodes are presented according to two different coding schemes. The first coding scheme (top two graphs) was used to count the frequency of scientific term use and the second coding scheme identified reasoning strategies employed by Brie and Mona (bottom two graphs). The two graphs shown in set (a) are from early in the unit where I claim that Brie and Mona were engaged in communication-oriented ritual routines. Set (b) illustrates discourse that took place approximately 160 minutes later (not the same day).

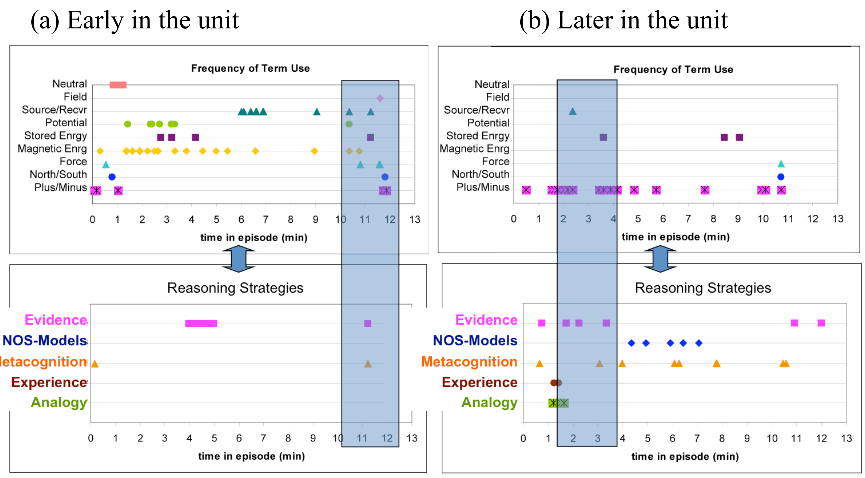



Figure 2. Comparison of students' use of scientific terms (top) and reasoning strategies (bottom). Set (a) is early in the unit and (b) is later in the unit.

As is evident from the graphs, the discourse that took place early in the unit—set (a) was dense with scientific terms such as charge, negative, positive, neutral, force, source, receiver, stored energy, potential energy and the same transcript is sparsely coded with reasoning strategies. Set (b) on the other hand, has fewer terms identified on the top graph and is rich with reasoning strategies as shown on the bottom graph. This suggests that the nature of the discourse was different in these two clips and supports the claim that focus on verbal, term-based communication was more the focus of early discourse and reasoning was more the focus of later discourse.

## Discussion

Like the early math learners in Sfard and Lavie's (2005) study, Brie and Mona practiced model-building in science initially in a largely communicative style, looking for appropriate terminology with which to communicate what everyone already knew. As the unit went on, Brie and Mona increasingly focused on the mechanism that could drive observable phenomena and less on communication for its own sake.

This seeks to demonstrate how discourse becomes objectified as a result of being practiced, and this is what defines the development of disciplinary knowledge. A reasonable question to ask is, "Why do students even have to go through the interpersonal discourse? Couldn't the teacher have just started with the iron filing experiment, telling students to focus on the entities inside the nail from the get go?" The point is not that the curriculum was explicitly set up to allow the students time to engage in interpersonal discourse. Instead, this was direction taken by the students' interaction with it. The unit on magnetism was particularly suited for this type of study because it consisted of a practice that Brie and Mona had little experience with—model building. Therefore, we were able to catch a glimpse at what it looked like as they embarked on early participation in this practice. The data demonstrate that the students' models of magnetism evolved toward the scientific model, their reasoning became increasingly mechanistic, and their discourse about the magnetic phenomena moved from the interpersonal domain into the domain of invisible entities that serve as a mechanism for observable behavior. As hypothesized by Vygotsky and further elaborated by Sfard and Lavie (2005) and countless others, learning involves participation. Participation involves the self and in order for the self to engage with the endorsed narrative of a discipline, she must access any available routine that is handy. In most of our cases, this is the routine of communicative ritual. We play this routine over and over as a means by which to catch on to the rules of play within that community. This is exactly what Brie and Mona did. They played communication as they investigated the problem space. In doing so, they caught on to subtle but critical practices that define success in mechanistic reasoning.

In this study, the task of creating a model of magnetism can be thought of as existing in two stages: (1) the task of creating product that was perceived as acceptable for the purpose of communicating with other individuals (term-dense sentences that could only account for the final conditions of the phenomenon) and (3) the task of objectifying the discourse so that the students' relationship with the objectified world of magnetic phenomena, and therefore the community of scientists, was able to flourish. The students began their model-building experience in discourse that centered on constructing term-dense sentences that could be presented to the others. Brie and Mona said things such as, "You used a word, what was it, Poten- Potential! Yea, I like that," indicating that they were looking for "good words" to use. They ended their model-building experience with a solid model of magnetic phenomena that could account for all of their observations. The entire class ended the unit with a discussion about the nature of models and their role within the scientists' activities. There is ample evidence from this study that the learning of disciplinary knowledge resulted from the students objectifying their discourse away from the self and its communication with other people and toward the self in relation to the human-independent world of magnetic phenomena.

## Acknowledgements

I am grateful to Ian HerManyHorses and Stephanie Jalovec who spent a year analyzing data with me in the beginning stages of this work.

## References

Abd-El-Khalick, F. (2001). Embedding nature of science instruction in preservice elementary science courses: Abandoning Scientism, but … *Journal of Science Teacher Education, 12*(3), 215-233.

diSessa, A. A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Meta-representational expertise in children. *Journal of Mathematical Behavior, 10*(2), 117-160.

Gee, J. P, Allen, A., & Clinton, K. (2001). Language, class, and identity: Teenagers fashioning themselves through language. *Linguistics and Education, 12*(2), 175–194.

Gee, J. P. (1990).*Social linguistics and literacies: Ideology in discourses.* New York: Falmer.

Gilbert, J. K., & Boutler, C. (1998). Learning science through models and modeling. In B. Fraser & K. Tobin (Eds.), *International handbook of science education* (pp. 53–66). London: Kluwer.

Goldberg, F., Robinson, S., & Otero, V. (2006). *Physics and everyday thinking.* Armonk, NY: It's About Time.

Griffiths, A., Preston, K., Harrison, A. G., & Treagust, D. F. (1998). Modeling in science lessons: Are there better ways to learn with models? *School Science and Mathematics, 98*(8), 420–429.

Gutiérrez, K. D., Baquedano-López, P., & Turner, M. G., (1997). Pulling language back into the language arts: When the radical middle meets the third space, *Language Arts, 74*(5), 368-378.

Gutiérrez, K. D., Baquedano-López, P., Alvarez, H. H., & Chiu, M. M. (1999). Building a culture of collaboration through hybrid language practices. *Theory Into Practice*, *38*(2), 87-92.

Harrison, A. G., & Treagust, D. F. (2000). Learning about atoms, molecules, and chemical bonds: A case study of multiple-model use in grade 11 chemistry. *Science Education, 84*(3), 352–381.

Hennessey, M. G. (1999). *Probing the dimensions of metacognition: Implications for conceptions of teaching-learning*. Paper presented at the National Association for Research in Science Teaching, Boston, MA.

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English language backgrounds. *Educational Researcher, 27*(4), 12-21.

Lemke, J. L. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex.

Machamer, P. (2004). Activities and Causation: The Metaphysics and Epistemology of Mechanisms. International Studies in the Philosophy of Science, *18*(1), 27-39

Moje, E. B., Collazo, T., Carrillo, R., Marx, R. W. (2001). "Maestro, what is 'quality'?": Language, literacy, and discourse in project-based science. *Journal of Research in Science Teaching*, *38*(4), 469-498.

Otero, V., & Nathan, M. (2008). Pre-service elementary teachers' conceptions of their students' prior knowledge. *Journal of Research in Science Teaching*, *45* (4), 497-523.

Roth, W.-M., & Duit, R. (2003). Emergence, flexibility, and stabilization of language in a physics classroom. *Journal of Research in Science Teaching, 40,* 869-897.

Schwarz, C., & White, B. (2005). Meta-modeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction,* 23(2), 165-205.

Sfard, A. & Lavie, I. (2005). Why Cannot Children See as the Same What Grown-Ups Cannot See as Different? – Early Numerical Thinking Revisited. *Cognition and Instruction,* 23:2, 237-309.

Vygotsky, L. S. (1986). *Thought and language,* Cambridge, MA: MIT Press.

Windschitl, M. (2004). Folk theories of "inquiry": How preservice teachers reproduce the discourse and practices of an atheoretical scientific method. *Journal of Research in Science Teaching 41*(5), 481-512.

# "Getting Others' Perspectives": A Case Study of Creative Writing Environments and Mentorship

Alecia Marie Magnifico, University of Wisconsin-Madison, 1025 W. Johnson St. Madison, WI, 53706
ammagnifico@wisc.edu

**Abstract:** Giving students opportunities to interact with real readers of their work may not only motivate them to write, but also to take on new literacies and see themselves as writers in new ways. I detail two case studies of successful writing communities—a high school classroom and an extracurricular arts program—and describe adolescent writing practices in the active (and often interactive) presence of the two different collaborative audiences. I discuss structural implications for the structure of authentic writing and writing workshop environments, the role of mentors in such instructional spaces, and the importance of teaching students how to be effective, collaborative audience members and readers of each others' writing.

## Introduction and Significance

In recent years, writing has become the new "basic literacy skill," as central as reading both in the workplace and in education (Brandt, 2001, 2009). Students write in many genres, both for school and for personal exploration and communication (Lenhart, Arafeh, Smith, & Macgill, 2008; Lunsford, et al., 2008). In schools, many states have mandated writing assessments, pushing teachers to emphasize writing throughout the curriculum (e.g. United States Department of Education, 2007). Only 25% of 12th grade students in United States schools were graded as "proficient" writers on the 2007 NAEP writing assessment (NAEP, 2008), however, suggesting a disconnect among the skill of writing, its real-world importance, and the teaching of writing in secondary schools. Part of this shortfall may be a result of the typical presentation of writing in secondary schools: a way to evaluate student knowledge, rather than a useful skill or a communicative tool (Applebee, 1996; Boscolo & Hidi, 2007; Nystrand, 1997).

While creative writing, the major form of in-school communicative and expressive writing, appears often in elementary schools, it drops out of most curricula by high school (Boscolo & Hidi, 2007). I argue in this study that mentors and teachers can encourage more authentic, communicative writing practice through design: Creative, expressive writing communities can motivate students to practice their own writing, to read others' writing, and to improve their broader literacies—all skills that prepare them to be writers and communicators in school and beyond.

This study explores relationships among instructional practices, instructional spaces, authentic writing, and adolescents' engagement in and understanding of creative writing. I take up two specific research questions: (1) At a design level, how are creative writing communities structured? What are the salient participant structures in which students and mentors participate? (2) What are the roles that students and mentors play? How do the structures affect these roles, and, more importantly, students' experiences and engagement with creative writing?

## Theory

In traditional classrooms, influenced by the belief that writing is more a tool for evaluation than an independent content area, most writing activity involves answering questions with a "correct" answer known by a teacher (Applebee, 1996; Nystrand, 1997). As a result, most classroom writing at the secondary level assesses content knowledge rather than students' ability to make an original argument, write cohesive narrative, or impact an authentic audience (Boscolo & Hidi, 2007; Cohen & Riel, 1989; Flower & Hayes, 1980). This particular focus, reinforced by state-level assessments as well as the number of subjects taught in schools, leads to a narrow view of writing instruction: content knowledge, structure, and Standard English conventions (Hillocks, 2002).

In contrast, students' writing in various authentic "communities of practice" (Lave & Wenger, 1991; Wenger, 1998), or "affinity spaces" (Gee, 2004; Lankshear & Knobel, 2007) is a tool for communication, creativity, and expression. These communities are found informally in places like online blogs and formally in publications such as literary magazines or club newsletters (Brandt, 2001). Differently from classrooms, the writing practice and instruction in these spaces tends to be "integrated" with preparing writing for an audience (Langer, 2001). Instead of completing a school assignment for a teacher, writers here must consider the communicative context of their work (Cohen & Riel, 1989; Fishman, Lunsford, McGregor, & Otuteye, 2005; Lunsford, et al., 2008).

Recent research into programs that bring authentic writing into classroom contexts has reported positive results, showing strong student engagement in the learning necessary to complete writing (Heath, 1983; Purcell-Gates, Duke, & Martineau, 2007), as well as critical thought about the audience and presentation of specific pieces of writing (Cohen & Riel, 1989; Dyson, 1997; Purcell-Gates, et al, 2007). In addition, work from writing and media studies suggests that young writers take on challenging topics and engage in key identity development when they have the opportunity to write for audiences in these ways (Dyson, 1997; Heath, 1983; Halverson, 2005; Hull, 2005).

Literacies research in secondary schools has established the writing workshop model—one that brings many features of authentic writing and affinity spaces into the classroom—as an effective practice for teachers (Alvermann, 2001; Atwell, 1998). Different from evaluative writing, workshop writing encourages students to think through ideas using writing as a tool, to read critically, and to understand classroom writing as preparation for advanced literacies (Alvermann, 2001). At the same time, both students and teachers need extensive practice with workshop techniques to be "effective responders" to each others' work (Simmons, 2003).

## Methods

I collected multiple forms of data from two creative writing environments in order to build distinct *instrumental case studies* (Stake, 1995) of successful environments in which the creative process, writing process, teaching and learning methods, and available audiences for the writing are different. These case studies exemplify two settings in which young people learn to write and practice writing: a school classroom and an extracurricular arts program.

### Participants, Settings, and Data Collection

The first case, a *school environment*, follows nine students (six female, three male) through a creative writing unit in an 11th grade English classroom at a K-12 suburban school. Two of the students identified as Asian exchange students, one as African-American, and six as White/Caucasian. The classroom teacher aimed for students to understand the genres of fiction and poetry better by working in small groups to write, revise, and analyze their own poems and short stories. These observations totaled approximately 15 hours over ten weeks.

The second case, an *extracurricular environment*, follows seven participants (five female, two male) at an urban creative writing camp for 9th-12th graders. Organized and staffed by local professional writers ("writing coaches"), the camp helps young people to hone their writing craft. Activities focused on daily writing, reviewing writing in small groups, and revising prose and poetry pieces for a final public reading. These observations totaled approximately 25 hours over one week.

I collected data from a variety of sources in both cases described above. In each setting, I observed and participated in one full *production cycle's* worth of work, which included the duration of creative writing instruction planned by the mentor(s). Despite the differences in each setting, the activities in which the participants engaged (e.g. writing, feedback, and revision) were comparable. Observations were captured in field notes and audio recordings. In addition, I conducted semi-structured interviews. I interviewed students twice, using pre- and post-interviews that were largely reflective, allowing them to speak in depth about their own writing processes, the pieces of writing that they worked on, and their experiences of class or camp. I interviewed mentors once, at the end of the writing instruction, about design choices that they made and their evaluation of the program.

### Data Analysis

Especially when considering a less-researched context such as secondary-level creative writing, qualitative methodologies such as open coding (drawn from grounded theory, e.g. Charmaz, 2000) and thematic analysis (e.g. Aronson, 1994; Braun & Clarke, 2006) are particularly important. While I have entered the analysis with specific hypotheses, these open-ended techniques allow me to identify, describe, and understand broad patterns in adolescents' perceptions of their online writing and the effects of these experiences on their learning.

This paper presents a preliminary thematic analysis (Aronson, 1994; Braun & Clarke, 2006) of these data. To conduct this analysis, I broke student interviews and classroom conversations into excerpts by question (for the interviews) or episode (for the field notes and classroom conversations). I began with open coding and refined my coding scheme in subsequent rounds of coding, highlighting central themes inductively from the data (e.g. "mentorship," "useful feedback") and deductively from the theoretical framework, (e.g. "affinity space structure," "authentic audience," and "evaluation").

### Pseudonyms — Student and Mentor Numbers

In order to track students and mentors across two cases, I have devised the following numbering scheme: All students in the *school case* are labeled with an "S#": In other words, they are identified as S1, S2, etc. All students in the *extracurricular case* are labeled with an "E#": E1, E2, etc. Since there was only one *teacher* in the school case, he is labeled "T1." The four *writing coaches* in the extracurricular case are labeled with a "C#": C1, C2, etc.

## Results: Major Findings

I present three themes drawn from observational data and student interviews, each examining an element of the design of the two instructional spaces and the effects of these design choices on the students. These themes are: (1) feedback structure, (2) role of mentors and teachers, and (3) how students interact with audience.

## Theme 1a: Feedback Structure — School

In the school case, observations demonstrated that class members (students and teacher working together) established a loose workshop structure among themselves. Students completed a short story and a small poetry collection (the teacher suggested 4-5 poems) during the ten weeks, but were able to choose which genre to take up first. Students were expected to bring writing to class on each workshop day, although not all students did so.

Students who had prepared work formed small groups during each class, read each other's work silently, made notes, and discussed their feedback with little intervention from their teacher, who provided assistance if requested. Students who were not prepared to share written work wrote silently during the class time. As a result of this structure, students received written and verbal feedback from their peers, usually the partners with whom they worked each day. Sometimes these groups remained consistent throughout the duration of one piece—for example, one pair of students worked together six times, repeatedly reviewing one poem and one story (field notes: 3/9/09, 3/20/09, 3/27/09, 4/1/09, 4/3/09, 4/14/09)—but often, students chose different partners during each workshop.

Students could (and a few did) consult with the classroom teacher directly for additional formative feedback. For example, S2 commented on his teacher's ability to help him focus on sensory details in his poetry: "I really like to focus on the big things and... it gets harder and harder for me to focus on actually expressing those thoughts in the poem, so he really pushed me to get those, the specifics." More formal written teacher feedback came in the form of a grade and comments after students had turned in an assignment.

## Theme 1b: Feedback Structure — Extracurricular

In the extracurricular case, observations demonstrated that each writing coach led a "writers' circle" of 6 students, which provided the central mechanism for students' sharing and getting feedback on their writing. The writers' circles met twice a day with consistent membership—each student received daily formative feedback from the same peers and writing coach. Some students worked on the same piece of writing throughout the week, bringing revisions or new sections to each writers' circle, while others brought several different pieces during the week.

The structure of writers' circle was proscribed and consistent throughout the week. Approximately 3 students shared their work in each writing circle. With two circles each day, each student was expected, though not forced, to share writing at least once a day. During each circle, one student read her work aloud while the others listened and made notes. After this reading, each member (including the coach), shared her feedback and passed her notes to the initial reader/writer. A short discussion occasionally followed, usually if many students gave feedback on the same aspect of a piece. For example, when two writers were struggling to resolve stories of loss and guilt., C1 began a conversation encouraging students to describe how they felt in these situations to help the writers capture these feelings realistically in their characterizations and descriptions (field notes, 8/5/09).

In the writers' circles, coaches listened and gave feedback, and students were often impressed by the details that coaches noticed: "[Coaches] have schooling and stuff on this topic... There's some things [my coach] noticed... [she saw] I was using adjective noun a lot, like that pattern. That was really good" (E6). If a student needed additional feedback, writing coaches also encouraged conferences during silent writing time (field notes, 8/5/09).

## Theme 1c: Feedback Structure — Looking Across Cases

The structure of these two learning environments may be represented in the following way:

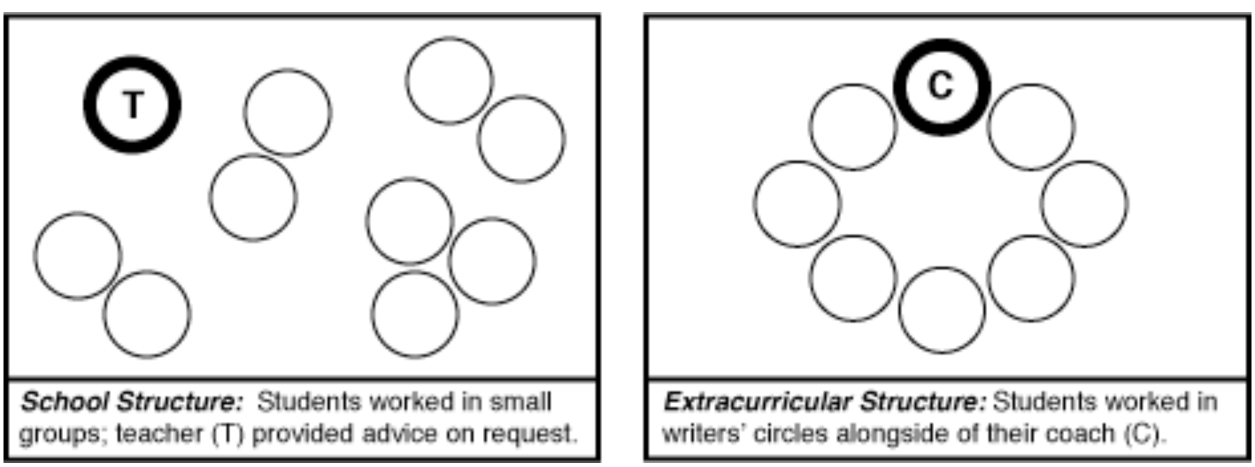



Figure 1: Structure of Learning Environments

All of the students in both cases (9/9 school, 7/7 extracurricular) recognized at least some value in getting feedback from readers (peers, mentors, or both) during the process of writing and editing, from "motivation... because I'm going to have to be sharing this" (E6), to "support... when I was scared to write poetry" (S4), to "other people's perspectives" (E5). In their interviews, students reported using the edits from the workshop groups and writers' circles—their peers and mentors—to improve their poems and short stories:

> It was helpful to see what other people understood... [One person] read this and said 'well, I'm not sure what you mean...' And then listed out potential ideas that they had... I saw what they took from it and could see... I need[ed] to emphasize these parts better (S6).

> Well, you don't have to follow [the feedback] but it's always good to know what other people thought... because if this part's not clear, or you can't see what this [character's] thinking, or it's not realistic that [a character would] feel this way after this happened, then it just opens your eyes. And then you work on it. And you get better by working (E3).

When audience members became collaborative editors, students in both of the case studies became invested in receiving criticism about the quality of and meanings behind their writings, and in working to revise with that feedback in mind.

## Theme 2a: Role of Mentors — School

In the school case, likely as a result of the design of the instructional space, observations and interviews revealed that the teacher did not have a clear role in mentoring the day-to-day happenings of the creative writing workshop sessions. He gave advice to individuals or small groups and kept time (and, occasionally, order) as students workshopped their writing, but he was not directly involved in the feedback structures of his classroom's workshops. He sometimes seemed unsure of his role in the new space that he and his students had built, although he was certain that he "d[id]n't want to grade them on their creativity" (T1), and spoke in class about not interfering with their ideas or process, "as long as [they were] writing" (T1). On the first day of workshop writing, just as students had begun reading each others' work, he came over to ask me "what do I do now?" (field notes 3/9/09).

In their post-interviews, a third of the students (3/9) expressed frustration with their teacher's role, specifically because it did not allow them access to his expertise unless they sought out his advice. Students appreciated that their teacher was trying not to apply "pressure... [like] whether or not you're [writing] what the teacher wants you to get" (S8) but at the same time, noted that his "adult opinion" (S8) was valuable to them. Many were eager to hear what he had to say about their writing: "[I wanted] to hear more of what he had to say a lot of the time... I think he [didn't give more feedback] maybe for the purpose of not having any influence" (S6).

## Theme 2b: Role of Mentors — Extracurricular

In the extracurricular case, the writing coaches were all familiar with the writers' circle feedback structure because they had used this design in prior writing camps (field notes, 8/3/09). Observations of and interviews with writing coaches demonstrated that all of the coaches (4/4) believed their central role to be supporting the students' writing by giving them "sacred writing time" to write and "good feedback on their work," through writers' circles. Almost all (6/7) of the students, too, connected the feedback from their coaches with improvement in their writing skills:

> [My coach gives] professional feedback. I mean, it's good across the board, the feedback [from my coach] that comes from experience, and then, from my fellow young writers, you get that feedback that just comes from the gut. [They tell me] I read this, and it kinda tripped me up. And that's kinda my audience... they will tell [me] what's right, what's good and what's not good (E7).

Students saw this feedback as vital to their development because they saw their coaches as an audience: experienced writers who knew about such difficulties as continuity, structure, and revising stories and poems for publication.

## Theme 3a: Writing for an Audience — School

While all of the students in the school case recognized the *value* of getting feedback from readers (peers, mentors, or both) during the process of writing and editing, they did not necessarily enjoy sharing their writing. The workshop process was new for almost all of them; 7/9 reported never having workshopped creative writing before. In their post-interviews, only one student in the class, S1, believed that getting feedback from her peers and teacher was always welcome and always useful, because it "can help you figure out if you got your meaning [across]." 4/9 students, however, noted that while they thought that getting feedback was helpful and led to improvements in their writing, they did not enjoy it because the process was often "uncomfortable" (S4, S6) or "too emotional" (S5):

> [In the workshop sessions] I guess I was thinking, oh they're not gonna understand what I'm trying to say here... I wanted them to get the meaning I got from [the poem], but sometimes I didn't want to share that emotion so... it was kind of like I [didn't] know what to do (S5).

One student in this group, S4 only showed her writing to the teacher for the entire 10 weeks. An additional 4/9 students noted that while they generally enjoyed getting and giving feedback, it was acknowledged that this process sometimes made them feel "worried" (S3) or "sensitive" (S8), especially when sharing a new piece for the first time:

> Yes [I liked getting feedback] for certain poems... I'd have to actually really feel good about them myself. I guess it goes with just about anything, you have to feel comfortable yourself with your creation before you show it to other people (S2).

While all of the students felt as if writing for an active, communicative audience of their peers was helpful in some ways, the process of doing so was very difficult for many of them. Particularly students who chose to write about personal, emotional topics felt as if their writing could expose "too much" of themselves (S9), and potentially parts of their lives that they did not know if they wanted to share with their classmates and teacher.

### Theme 3b: Writing for an Audience — Extracurricular

In the extracurricular case, students reported more positive reviews of the process of giving and receiving feedback , particularly in the context of writers' circles. This finding is unsurprising in many ways, since all of the students in this case (7/7) had been to this creative writing camp in prior years, and were therefore familiar with the process of sharing their writing with their peers and writing coach. In their post-interviews, all of the students (7/7) reported valuing the feedback that they got, while 6/7 students noted that they enjoyed sharing their writings:

> I really liked sharing in the writer's circle. I think that we had a really good group too, because they, they evaluated a lot... I dunno, they helped. They didn't just say I liked it, they were good critiquers. [T]hey gave a lotta criticism, but also praise, and... they almost talked about the idea that you had, and what they thought it was. It helped me [to see] other people's perspectives (E5).

A few (3/7) students noted some apprehension, reporting that they felt "kind of anxious" (E1), or "nervous" when sharing a new story (E2), but all of the extracurricular students tempered these feelings with stories of how peers and coaches helped them to improve their writing, or inspired them to keep going with a difficult piece of writing.

One unexpected finding was that many students (4/7) in the extracurricular case did not only talk about why they enjoyed camp, but about why they did not like sharing their writing at school. Their main complaint was the lack of constructive criticism, or, as E1 stated, "they just are like ohmigod, this is really good... how am I going to improve off of that?" The biggest condemnation of sharing outside of camp, however, came from S6, who argued that English teachers should "never" force students to read creative work aloud in school:

> [F]or school, I hate it when they have us read stuff out loud... [B]ecause all of the ideas I had that would've been perfectly fine at camp... they wouldn't be the same if they were told to all of my classmates. They would just think of the idea as really weird, instead of just appreciating it, and so then I couldn't think of anything [to write]... I knew that they were all gonna be reading me and judging me on it because you know, it's sort of a school environment where everyone judges you on what they hear...it wasn't like that at all at [camp] (S6).

Whereas sharing her writing with camp students—other writers who appreciate unusual ideas—is easy, sharing her writing at school is difficult. While the collaborative audience of her writers circle at camp would help her improve a "really weird" idea for a story, the evaluative audience of her classmates would judge her for her creativity.

## Discussion and Implications

In both of these cases, the mentors made efforts to reduce the differences between themselves and their students by avoiding many evaluation structures that are typical to schools and working to create a space in which students felt free to pursue creative, productive writing. Instead of commenting summatively on students' work, mentors in both cases made themselves available to give formative feedback and allowed students to choose topics freely. While students in the school case had some limits—the necessity of exploring both prose and poetic genres, an analysis of their own work—students in the extracurricular case students explored writing as they wished. Largely, these design elements helped both of these spaces to cohere as successful writing environments. In the school case and in the extracurricular case, students wrote several creative pieces, successfully solicited and received feedback from their peers, mentors, or both, and revised their writing to the specifications of that feedback. (All students in both cases revised at least one piece of writing: Reports on revision data are forthcoming.)

It is worthwhile to note, however, that the purpose of writing in each of these environments is quite distinct, and affords feedback and collaboration differently: As many sociocultural theorists note, context matters (e.g. Black, 2008; Dyson, 1997; Heath, 1983). Students who engage in creative writing as part of a school program are forced to do so, at least at some level. While many students did explore themselves, their emotions, and their histories as topics for their poems and prose, it was difficult for some to engage in this important identity work at school. This kind of writing exposed sensitive feelings in front of classmates and for a grade—a situation where risk-taking is understandably dangerous (Dyson, 1997; Halverson, 2005). In contrast, the students in the extracurricular case had much more freedom to explore "weird" writing at camp (as S6 noted), partly because other creative writers are tolerant of offbeat ideas, but perhaps also partly because camp is necessarily separate from everyday communities.

Combining the ideas of purpose and mentor roles in these two case studies sheds light on the importance and the difficulty of establishing *truly* collaborative audiences in a writing workshop environment. Not only must writing instructors design a space that supports purposeful, expressive writing and establishes a structure for students' collaborative interactions, but a structure for mentors' participation as well.

In the school case, the teacher required that his students write and he stood outside of the collaborative feedback structure. While the intention of this design choice was to allow greater expression within a required assignment—he spoke often of "not grading students' creativity"—this move backfired. Because he was outside of the collaboration, teacher feedback came only when students sought out his advice or had turned in an assignment, and thus was seen as summative. Despite the workshop-oriented structure of the classroom, and the students' collaborations, it was difficult for the teacher to avoid the norms that are inherent to classroom communication and evaluation (Applebee, 1996; Nystrand, 1997). In the end, it seems that the power structure of school cast him as an evaluator, although another result may have been possible if he had participated more fully in the workshop.

In contrast, writing coaches were integral to the feedback structure in the extracurricular case, acting as expert writers and representatives of a community of practice that the participating students sought to join (Lave & Wenger, 1991; Wenger, 1998). Students chose to attend writing camp, many of them for several years running, because they appreciated this kind of expert instruction. Coaches modeled listening and feedback in writers' circles, and often drew connections across young writers' experiences, pointing out broad patterns in students' writing and offering ideas for how they might achieve an attempted technique. Instead of their role as leaders forcing them into evaluation, the writing coaches used their place inside of the writers' circles to encourage debate, frame students' comments, provide additional insights, and offer questions or future directions for students' writing. This was, however, no doubt easier for coaches who did not have to evaluate students in the end—this structure likely made it much easier for students to see their feedback as collaborative and formative.

One interesting reading of this difference is in the separation between the goals of the school environment and that of the extracurricular environment. In the extracurricular case, the coaches saw their role quite literally as coaches: Their job, in the week-long camp, was to help young, inexperienced writers improve as much as possible, which requires intensive modeling and guidance from listeners with writing experience (Simmons, 2003). In contrast, the teacher in the school case was not invested in modeling writing, but in his students' learning of literary devices through experimentation. He wanted his class to learn how to think about why writers make certain choices, connecting their own struggles with figures of speech (for example) with the analysis of stories and poetry written by more famous authors—common practices in secondary English classrooms (Applebee, 1996).

While the creation of workshop groups with collaborative, authentic audiences was similar across both cases, the purposes of writing and the overarching instructional structures were very different—and this difference had a profound effect on the mentors' roles and the students' experiences. Can students and mentors participating in instructional spaces serve as authentic readers and audience members for each others' writing? This preliminary analysis suggests that these kinds of interactions are possible, particularly with practice and a carefully designed structure. At the same time, the transformation from traditional classroom to authentic writing workshop is not a straightforward one for teachers or students, and should be approached with deliberation and attention.

## References

Alvermann, D. (2001). Effective literacy instruction for adolescents: Executive summary and paper commissioned by the National Reading Conference. Chicago, IL: National Reading Conference.

Applebee, A. (1996). *Curriculum as conversation: Transforming traditions of teaching and learning.* Chicago, IL: University of Chicago Press.

Aronson, J. (1994). A pragmatic view of thematic analysis. *The Qualitative Report, 2*(1). Retrieved 30 June, 2009, from http://www.nova.edu/ssss/QR/BackIssues/QR2-1/aronson.html

Atwell, N. (1998). In the middle: New understandings about writing, reading, and learning. Portsmouth, NH: Boynton/Cook Publishers, Inc.
Black, R. (2008). *Adolescents and online fan fiction*. New York: Peter Lang.
Boscolo, P. & Hidi, S. (2007). The multiple meanings of motivation to write. In P. Boscolo & S. Hidi (Eds.), Studies in Writing, Vol. 19, *Writing and motivation* (pp. 1-14). Oxford: Elsevier.
Brandt, D. (2009). *Literacy and learning: Reflections on writing, reading, and society.* San Francisco, CA: Jossey-Bass.
Brandt, D. (2001). *Literacy in American lives.* Cambridge, UK: Cambridge University Press.
Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77-101.
Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In Denzin, N.K. & Lincoln, Y. S. (Eds.) *Handbook of qualitative research, 2nd edition* (pp. 509-536). Thousand Oaks, CA: Sage.
Cohen, M. & Riel, M. (1989). The effect of distal audiences on students' writing. *American Educational Research Journal, 26*(2), 143-159.
Dyson, A.H. (1997). *Writing Superheroes: Contemporary childhood, popular culture and classroom literacy*. New York, NY: Teacher's College Press.
Fishman, J., Lunsford, A., McGregor, B. & Otuteye, M. (2005). Performing writing, performing literacy. *College Composition and Communication, 57*(2), 224-252.
Flower, L. & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication*, *31*(1), 21-32.
Gee, J.P. (2004). *What video games have to teach us about learning and literacy.* New York, NY: Palgrave-Macmillan.
Halverson, E.R. (2005). InsideOut: Facilitating Gay Youth Identity Development Through a Performance-Based Youth Organization. Identity: An International Journal of Theory and Research, 5(1), p. 67-90.
Heath, S.B. (1983). *Ways with words: Language, life, and work in communities and classrooms.* New York, NY: Cambridge University Press.
Hillocks, G. (2002). *The testing trap: How state writing assessments control learning*. New York, NY: Teachers College Press.
Hull, G. & Nelson, M. (2005). Locating the semiotic power of multimodality. *Written Communication, 22*(2), 224-261.
Langer, J.A. (2001). Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal, 38*(4), p. 837-880.
Lankshear, C. & Knobel, M. (2007). Sampling the "new" in new literacies. In Knobel, M. & Lankshear, C. (Eds.) *A new literacies sampler.* New York: Peter Lang.
Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, UK: Cambridge University Press.
Lenhart, A., Arafeh, S. Smith, A., & Macgill, A. (2008). Writing, technology, and teens. In *PEW Internet and the American Life Project.* Retrieved 20 March 2009, from: http://www.pewinternet.org/Reports/2008/Writing-Technology-and-Teens.aspx
Lunsford, A., et. al. (2008). *Stanford Study of Writing.* Retrieved 2/15/2010 from: http://ssw.stanford.edu/research/research.php
National Assessment of Educational Progress (NAEP). (2008). 12th grade. In *The nation's report card: Writing 2007 Report Card.* Retrieved 15 September 2009, from: http://nationsreportcard.gov/writing_2007/
Nystrand, M., A. Gamoran, R. Kachur, & C. Prendergast (Eds.). (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom.* New York, NY: Teachers College Press.
Purcell-Gates, V., Duke, N. K., & Martineau, J. A. (2007) Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly, 42*(1) 8-45.
Simmons, J. (2003). Responders are taught, not born. *Journal of Adolescent & Adult Literacy, 46*(8), p. 684-693.
Stake, R. (1995). *The art of case study research.* Thousand Oaks, CA: Sage Publications.
United States Department of Education. (2007). Writing framework and specifications for the 2007 National Assessment of Educational Progress. In the *Institute of Education Sciences National Center for Education Evaluation and Regional Assistance*. Retrieved 15 September, 2009, from http://nces.ed.gov/nationsreportcard/writing/whatmeasure.asp
Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity.* Cambridge, UK: Cambridge University Press.

# Cross-disciplinary practice in engineering contexts – a developmental phenomenographical perspective

Robin Adams, Tiago Forin, Saranya Srinivasan, Purdue University, West Lafayette IN
Email: rsadams@purdue.edu, tforin@purdue.edu, sap@purdue.edu
Llewellyn Mann, Swinburne University of Technology, Melbourne AU, suneagle@hotmail.com

**Abstract**: This paper presents an overview of results from a phenomenographic study that investigates critical differences and similarities in the ways people experience cross-disciplinary practice in engineering contexts. Study implications are discussed regarding developmental perspectives on cross-disciplinary ways of thinking, acting, and being.

## Introduction

Many complex engineering problems facing society such as climate change, healthcare and international security require cross-disciplinary approaches that integrate diverse perspectives into a collective whole. Here, "cross-disciplinary" refers to practices associated with thinking and working across different perspectives such as multidisciplinary, interdisciplinary, and transdisciplinary. Reports on the future of engineering education stress preparing engineers to become "emerging professionals" who can deal with complexity, innovate, flexibly adapt to new situations, and bridge disciplinary boundaries to produce deeper insights (NAE, 2004). There has been substantial investment and interest in cross-disciplinary practice; however, the level of empirical attention is considerably less (Bromme, 2000). Studies focus on humanities and social sciences (e.g., Klein, 1996; Lattuca, 2001), although there are some studies on the physical and natural sciences (Galison, 1997; Nerssessian, 2006; Newstetter et al, 2004) and engineering design (Adams et al, 2009a).

While understanding how groups work together and create new knowledge is an important research focus, this study moves the lens from groups to individuals to make visible the development of cross-disciplinary practices. Phenomenographical methods were used to investigate critical similarities and differences in how individuals experience and comprehend cross-disciplinary practice in engineering contexts. Using experience as a lens for investigation acknowledges the relationship between experience and learning (Tuan, 2007, p. 16): "the modes through which a person knows and constructs a reality…and learn from what one has undergone." Phenomenography also provides unique understandings of practice that can readily contribute to course and program renewal and redevelopment (Daly et al, 2008). In this paper we present an overview of study results and discuss some implications for research and education.

## Theoretical Frameworks

Three frameworks guided this study. The first framework involved a synthesis of cross-disciplinarity at the group level of analysis (Adams et al, 2009a) that revealed important variations regarding: problem orientation (from thematic oriented projects to systems oriented participatory projects), mode of knowledge production (from juxtaposition of perspectives to an overarching and transformative synthesis), outcome (from no change in knowledge to knowledge fusion through critical reflection), interaction structures (from collaborating as disciplinarians to transcending disciplines), discourse practices (from common interests to creation of new language and logic), and impacts on participants (from retaining a disciplinary identity to critical reflection on pluralistic identities). While the synthesis illustrates how multidisciplinarity, interdisciplinarity, and transdisciplinarity differ in significant ways, they all presuppose that cross-disciplinary practice is driven by problems beyond the confines of any single discipline and involves transgressing disciplines. Disciplinary practices are systems of inquiry characterized by theories and concepts, and modes of observation, interpretation, and arguing claims regarding truth and validity (Klein, 2004). Cross-disciplinary practice at the individual level is likely to involve confronting, recognizing, and respecting different epistemological views (Lattuca, 2001; Klein, 1996) in which deep learning interacts with identity development (Sfard & Prusak, 2005). This synthesis aided this study by (1) highlighting important variations in cross-disciplinary practice to guide participant recruitment, (2) revealing the need for empirical research at the individual, not group, level unit of analysis, (3) indicating theories of epistemological and identity development that may emerge from this study, and (4) characterizing cross-disciplinary practice as a situated practice.

The second framework is based on Dall'Alba's (2009) model of professional development. Dall'Alba and Sandberg (2006) describe understanding as embedded in dynamic, intersubjective practice – a notion that integrates thinking, acting, and being into an unfolding "professional-way-of-being" that makes visible the situated nature of professional development, variations in skill development, and characteristics of skillful performance. This connects to theories of situated learning, reflective practice, and communities of practice. For this model, skill progression represents how individuals might become increasingly adept at tasks that they routinely practice. Skillful performance (an embodied understanding of practice) represents a limited number of qualitatively different ways a practice is understood. Advancement along this dimension is marked by increasingly comprehensive ways of understanding – an "unfolding circularity" of ongoing inquiry to achieve more complex and comprehensive understandings of the central ideas of their practice. This framework aided this study by (1) identifying ways to characterize an embodied understanding of cross-disciplinary practice, (2) further locating this study in theories of situated cognition, and (3) informing the choice of phenomenography as a research framework.

Phenomenography is a qualitative method to investigate different ways of experiencing and understanding aspects of practice (Bowden & Green, 2005; Marton & Booth, 1997). It is used to reveal surface and deep approaches to learning and has been used to investigate the experience of learning, teaching, concepts such as programming and velocity, and more recently professional practice. In phenomenography the "phenomenon" is a person's experience in relation to an aspect of the world and the outcome of phenomenographic study is the researcher's interpretation of that phenomenon (Mann et al, 2007). It is an empirically derived method where a theory of variations is emerging as a way to explain the existence of categories and how they are related. In phenomenography, the goal is to identify categories of description comprising distinct groupings of qualitatively different ways of experiencing a phenomenon and internal relationships among groupings as increasingly complex ways of experiencing the phenomenon and a growing awareness and comprehension of practice (Bowden & Green, 2005). In this way phenomenography creates a link between experience and awareness where variations in experiences may reveal characteristics of an embodied understanding of practice.

## Research Design

Twenty-two (22) engineers and non-engineers who work in engineering contexts were strategically recruited to maximize diversity and establish an inclusive "outcome space". This sample size and strategy is consistent with typical phenomenographic studies. Variations in the sample were identified through a literature review and include: context of work (academia, private industry, and community service), years of cross-disciplinary experience, gender, nature of experience in terms of project scale (i.e., size of teams) and complexity (i.e., number of disciplines involved), and epistemological distance (i.e., scale of disciplinary similarity or difference). Epistemological distance is the extent to which an individual interacts with those with similar ways of knowing such as an engineer working with other kinds of engineers or substantially different such as an engineer working with a social scientist or artist. Figure 1 shows key sample variations: years of experience, project scale, and epistemological distance.

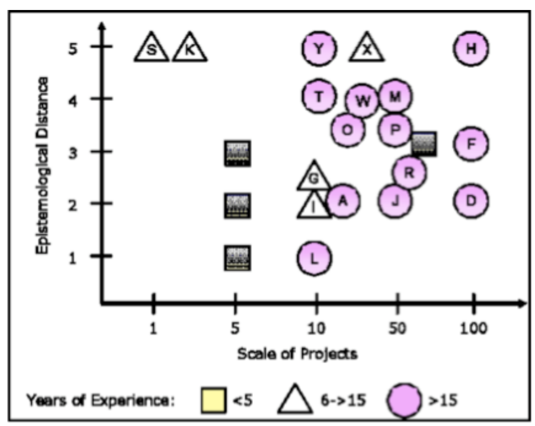



*Figure 1. Sample distribution among key characteristics. Letters represent participant identifier code (e.g., G = Grace, I = Isabella)*

Data was collected using a semi-structured interview protocol in which participants were asked to bring to mind concrete experiences of cross-disciplinary practice (see Mann et al, 2007). Follow up questions ask interviewees to elaborate on their experiences and what they mean by certain concepts. Probes may include information about the situation or characteristics of that situation that represent their ideas of cross-disciplinary practice. All probes were based upon what the participant expressed so far in the interview, and not formed through predetermined ideas and questions from the interviewer. Interviews lasted 30 minutes and were audio recorded then transcribed.

Analysis began with reading and re-reading transcripts as a collection and identifying what was significant within a whole transcript and in relation to other transcripts. Transcripts were sorted into piles of similarities and differences and assessed by identifying substantive excerpts from the transcripts that supported patterns in each pile. Over time the number and description of the different piles settled and became the categories of description. Subsequent passes through the data focused on articulating and substantiating the relationships among these categories, drawing from the data as well as theoretical frameworks. Relationships among categories were examined in terms of how categories related to a set of common ideas and to differences in breadth and depth of understanding. In this study, the relationship among categories mapped out increasingly more comprehensive understandings of cross-disciplinary practice. Overall, the analysis was a rigorously iterative process of being disciplined by the data, looking for empirical evidence of patterns, and seeking logical arguments in how patterns mapped to theoretical frameworks. In this way findings emerged through a co-evolutionary process of discovery of categories and construction of structural relationships between categories.

## Results

Below we summarize key findings: the four categories of description that emerged, critical variations across categories, and relationships among categories that reveal increasingly advanced ways of experiencing and comprehending cross-disciplinary practice in engineering contexts. Progression from Category 1 to Category 4 represents increasingly complex ways of experiencing and understanding cross-disciplinary practice in engineering contexts. In this paper we provide illustrative examples of the categories of description and present new findings on the hierarchical relationships among categories. An in-depth analysis of the categories is provided elsewhere (Adams et al, 2009b).

### Category 1: Working together – Emily, Grace, and Isabella

The experiences in Category 1 illustrate *cross-disciplinary practice as working together with people who have different training to effectively find a better solution*. The following excerpts highlight some of the key attributes of this category: (1) an iterative process of asking questions, challenging assumptions, and listening for understanding, (2) being comfortable with asking for information that might seem obvious, (3) knowing what you and others contribute, and (4) recognizing differences in what people know and how they communicate, (5) and the need to take personal responsibility to be an effective collaborator.

> "For me cross-disciplinary practice is probably working with someone of another discipline in a rather intricate way. In a way that's as brief as possible so that each party can continue their own work, but as thorough as possible when you get together so that you have very good information for each other…what needs to be improved, what details need to be in there." [Emily]

> "I mean it's very much learning how to talk to people that don't come from the same background that you do, academically. Being able to communicate what you want to do in a language that is accessible to an average level PhD scientist but not someone inside your own discipline necessarily….Just don't jump right in to the gritty details of it." [Isabella]

> "…probably the biggest challenge is getting the two to communicate…to listen and be very clear as to what you're trying to say and to make sure that the people listening to you are understanding that so, listening, asking questions to follow up is the biggest thing." [Grace]

> "I think it's vital. I think that's the only way to do it. If you don't have value for the other discipline, what are you even doing working with them? If you don't value what they are doing, you are not going to care how they do it, why they do it. You're not going to ask enough questions, and the questions are going to be important for the way you design what ever you're designing." [Emily]

> "Give people more respect as far as what they can contribute….I found that there were lots of good ideas that came from the engineers when they spoke up and there were lots of really important things that the nursing staff could contribute that I didn't think they could. Engineering people look at nursing people and think they're smart but in a whole different way that's not really that smart. They don't think of it as being difficult knowledge because it's all really, it seems like common knowledge like take chicken soup for a cold or something….really have their own special way of knowledge and I've come to appreciate that a lot." [Emily]

### Category 2: Intentional learning – Uri, Olivia, Pablo, Brianna, Ryan, Nadia

Where the experiences in Category 1 focus on the dynamics between people, the experiences that represent Category 2 focus on individuals' intentional learning. For Category 2, *cross-disciplinary practice is intentional learning so that everyone gains (me, my team, and my stakeholders).* The following excerpts highlight some attributes of this category: (1) creating opportunities to learn new perspectives or ways of knowing, (2) purposefully educating each other to collectively enable a systems perspective, (3) learning *through* experience and failure, (4) learning *how* to negotiate meanings across perspectives and formulate or investigate problems through multiple lenses, and (5) a passion and appreciate for continual learning.

> "…and every time I work with one of these people you have to understand something about what they do and you always, when you first start and everybody I've ever talked to seems to have this same perspective; you underestimate how hard the other person's field is. You've spent all this time and effort developing your capability; and oh, by the way you'll pick that up too and you learn pretty quickly if/that if you're not just going to do this superficially; that you know they've traveled their long road too. And so you have to meet as equals but you have to, and respect each other, but you have to learn about they're problem if you're really going to be successful in at least some amount and commit to that. But that's fun so…" [Olivia]

> "…a you help me/I help you kind of thing but and I would come back and I would never know enough so you'd have to read more and the best way of doing it was in an immersion kind of mode..." [Olivia]

> "I think I've become more convinced that the people need, it's like you go along in you need your immunizations, you need to have these points in life of injection of something and I think that it's important to plant those seeds along the way and I haven't really thought about that…" [Olivia]

> "I want to say I just became more open minded towards other disciplines. I used to think, we're engineering, it's much more difficult than everything else and although we can't communicate as well as the business people, but we're more technical, but when you start working with them, everyone sort of has their own set of abilities and weaknesses that they have to work together. I think by working cross-disciplinary, it sort of strengthens everybody." [Uri]

**Category 3: Strategic leadership – Jacob, Michael, Fergus, Tyler, Daniel, Yvonne**

Where the experiences of Category 2 focus on learning, the experiences that represent Category 3 focus on applying learning to actively enable cross-disciplinary work and outcomes. In other words, *cross-disciplinary practice is strategic leadership to enable cross-disciplinary work and synergy for the best outcome*. Leadership is central in that it involves being the "interface", "connector", or "communication specialist" to enable innovation or proactively transform a negative working environment into a positive one. The following excerpts highlight some attributes of this category: (1) make or enable conceptual connections, (2) build allegiances and trust, and (3) facilitate systems-oriented strategies or frameworks that leverage diverse perspectives. Some strategies involve actively transforming a negative working environment into a positive one.

> "…cross-disciplinary engineering is creating or innovating across multiple disciplines, so that's just saying it differently. But the strongest thing that does come for me is across markets or educational pillars. The other one which doesn't fit real nicely, but these days a lot of people do, is across different cultures." [Daniel]

> "I've got components, now I've got to think about, based on this organization, who are the right people that I need to speak with and get on a team so that we can ensure that we drive this thing forward." [Fergus]

> "I'm a communications specialist…I get a sense for people so I also get a sense of what they mean when they're telling me something. And so I would use that capability to get minds met. I would use a lot of other techniques, but I was a bridge." [Jacob]

> "…I believe, because there is a possibility that people who come in late don't get the same deal, don't get treated the same way. You've missed a lot of the discussions, you don't understand. In addition, if you're really adding a valuable team member, the discussions probably weren't as productive and fruitful as they could have been without his missing voice. So if you believe in a balanced approach to the problem-solving which I happen to believe, statistically you're more likely to have a better solution if you have more voices that are reasonable and respected sitting in the room….So balance, to me, says that you, you and I and three or four other people, whatever the size of the team is, are sitting around and we respect each other, ok, and we're willing to listen to each other and consider that these things, these issues that are raised or solutions that are raised are real and should be considered. That's what I mean by balanced problem solving….So everybody's in on the ground floor. It's important because you get a better solution in my opinion. You also have owners of the solution." [Jacob]

> "We're so busy trying to be sure that we're understood by others, that we don't seek to understand….So that is absolutely critical in order to be trusted, to trust people, you have to be trustworthy yourself. So it really starts with you and then you model behavior you want. It's also important in this team. And then probably the one that (captures) it best is believing in the value of synergy." [Jacob]

**Category 4: Challenging and transforming practice to integrate systems – Anthony, Xavier, Hannah, Logan, Samantha, Wendy, Kelvin**

Where the experiences of Category 3 focus on enabling cross-disciplinary work, the experiences of Category 4 represent transformative reflective practice that challenges prior training and ways of thinking.

For this category, *cross-disciplinary practice is challenging and transforming practice to integrate systems and produce an outcome greater than the sum of its parts*. The following excerpts highlight some attributes of this category: (1) critically challenging disciplinary practice and the ways conflict can be transformative, (2) integrating stakeholders as collaborators, (3) attuning to the human aspect of complex systems, (4) advocating perspectives by taking into account the broader context, and (5) embracing cross-disciplinarity as everyday practice.

> "Perspective really comes from the experiences that one has had and not one person has had the same experiences....that gives me a different perspective because I've seen different things, I've been exposed to different things. Maybe I've traveled all over the world, or no, I've only lived in Indianapolis all my life. You know those kinds of things; it gives you, that brings perspective…" [Hannah].

> "For better or worse I am who I am and some people appreciate that….I don't do anything different here than I do at home as I'm looking through something. So this is not my 'day' personality." [Logan]

> "I've always thought that it's always a good interaction from the standpoint of a multi-disciplinary scientist, that when you go in you work with the stakeholders you're at, you're on equal ground, you're on equal footing. There is an exchange of information. They're kind of framing the problem in a much different way than what you see it. You tend to be more isolated as a scientist, as someone who is at a university. They're dealing with this on a day-to-day basis, they have the muddy-boots kind of experiences. So the way in which they articulate the problem, they way in which they see it, the way in which they, the constraints that they have, is very, very valuable to listen to that….So, working with stakeholders is not just, you know, something we take lightly. It's actually part of the science. It's incorporated into the whole methodology." [Xavier]

> "…there are a whole bunch of metaphors, like the idea of poaching or borrowing or, I can't think of any more now, ok there really are a whole bunch of metaphors, that I don't see as transformative enough. That's great; you went out and borrowed an idea from somewhere else, and you're like so how can I use that tool in my stuff? But does that transform your stuff? Does that transform your discipline? I don't think so." [Samantha]

> "Well, you know, we did it collaborating with social scientists, right, in some respects. I mean, I've invested a tremendous amount of my own time trying to learn what these methods are. You know, I'm reading, you know, qualitative methods. I've got the books around here too that my colleagues read and so I'm making the personal investment to sit down and learn what I can about their practice of science and how it integrates with my traditional practice of science. How do you structure mine and how do you structure theirs to come and do, you know, make the interface work." [Xavier]

> "…if either of the two disciplinary scientists aren't willing to give up some turf, it's never going to happen, you know, it's never going to happen productively." [Xavier]

> "does the home need to have institutionalization or institutional markers or some kind of outward existence or can it be a group of people who identify in certain ways are, that don't find their cognitive home somewhere else? I don't know. But it's a more dynamic place. It is a more revolutionary place." [Samantha]

### Relationship Among Categories

The categories of description that emerged from the analysis are *distinct* – each captures a qualitatively different facet of cross-disciplinary practice in engineering contexts. These map to the framework for an embodied understanding of professional practice in terms of thinking (awareness of "difference", situation complexity, and goal direction), acting (engaging with "difference" and situation complexity), and being (self-perceived role or identity). Relationships among these categories are also *hierarchical* (see Figure 2). While attributes of each higher category encompass attributes of prior categories, critical variations of the more comprehensive categories are not evident in less comprehensive categories (see Table 1). This hierarchical progression is both data-driven (emerged from the analysis) and logical (attributes of higher categories move from surface to deep and narrow to broad comprehensions).

*Figure 2. Relationship among categories of variation*

Category 1 (Working Together) is distinct because of a focus on the experience of collaborating and communicating with people who have different perspectives, language, interaction styles, and ways of thinking. This appears to be a foundational category since the other categories build off of these ideas in increasingly complex ways. As shown in Table 1, critical attributes of Category 1 involve an awareness of differences in disciplinary training and how these differences complicate the process of working together towards an effective outcome. This awareness often triggers an iterative communication process of asking

TABLE 1. CRITICAL VARIATIONS IN CROSS-DISCIPLINARY PRACTICE

| | Category 1: Working together | Category 2: Intentional learning | Category 3: Strategic leadership | Category 4: Challenge & transform practice |
|---|---|---|---|---|
| AWARENESS OF COLLABORATION DIFFERENCES | • Difference = disciplinary training<br>• "Difference" makes collaboration complicated<br>• Everyone "smart" in own way<br>• Respect as "value-added" | • Difference = disciplinary perspective<br>• "Difference" as opportunity<br>• Respect for difficulty of disciplinary training | • Difference = disciplinary, organizational, cultural<br>• "Differences" clash; need strategies and new paradigms to enable collaboration and innovation<br>• Aware of need for common ground, trust, shared ownership<br>• Respect for the interfaces | • Difference = lived experiences; transcends disciplinary boundaries<br>• "Differences" enable transformation<br>• Honor other ways of knowing; no way is better than another |
| AWARENESS OF SITUATION COMPLEXITY | • Complexity as function of client and technological needs<br>• Aware of limits regarding technological feasibility<br>• Problems as client-driven, theme-driven | • Complexity as function of societal and global dimensions<br>• Aware of limits of own knowledge<br>• Problems as emergent | • Complexity drives need new paradigm of societal, global, economic views<br>• Aware of limits of disciplinary approaches and frameworks<br>• Problems require organizational thinking | • Complexity requires attuning to human context<br>• Aware of stakeholder risks and consequences<br>• Aware of system boundaries (disciplinary, organizational, cultural)<br>• Problems require integrated technological – social – human systems thinking |
| AWARENESS OF GOAL | • Meet needs, expand application, identify what is feasible from another disciplinary perspective | • Learning as its own goal – self, team, stakeholders<br>• Interaction yields new insights<br>• New partnerships | • Worthwhile outcome for all<br>• Successful launch<br>• Common ground as an outcome | • System level and participatory outcome greater than sum of parts<br>• Shape decisions<br>• Give up disciplinary turf<br>• Integrative synthesis that transcends disciplines<br>• Transform views on good practice, good science |
| ENGAGEMENT WITH SITUATION | • Iterative communication process | • Self-directed intentional learning process<br>• Create opportunities to learn - immersion and failure<br>• Build partnerships | • Change the parameters to discovery and cross-disciplinary paradigms<br>• Focus on problem formulation through multiple perspectives (not "jump too quick to implementation")<br>• Create collaboration networks | • Work at the interface; work at mile high system view<br>• Complex systems thinking<br>• Bring all aspects of lived experiences together<br>• Passion to improve system performance |
| ENGAGEMENT WITH DIFFERENCES | • Client / partner as information source<br>• Take responsibility for being effective communicator | • Meet as learners, educators<br>• Learn how to think between disciplines, synthesize<br>• Learn how to collaborate | • Proactively manage and leverage differences<br>• Orchestrate discovery and innovation<br>• Bring people together around shared goal; build expert network<br>• Build trust, shared ownership, allegiance, common ground | • Engage stakeholders as partners<br>• Advocate different perspectives; be altruistic<br>• Diversity and conflict as transformative |
| IDENTITY | • Responsible for contributing training as disciplinary team member<br>• Team member | • Passion for learning, dealing with complexity<br>• Learner and educator | • "Let up on the ego" – enable team<br>• Take the risk and lead<br>• Facilitate synergy at interface, synergy angels<br>• Interface, bridge, translator, etc. | • Live at interface as transformational learning site<br>• Revolutionary new "home"; perceived loss of disciplinary home<br>• Cross-disciplinarity as everyday practice – integrates work and life |

From Category 2 to 1: Improving the conditions needed to work together with people with different training to address complex problems of social and global significance

From Category 3 to 2: Applying prior learning on the challenges and affordances of cross-disciplinary practice to proactively enable successful cross-disciplinary discovery and innovation

From Category 4 to 3: Critical reflective practice to enable transformative learning and outcomes (for self, teams, stakeholders, and disciplinary practices) questions and listening for understanding with those who are perceived as a relevant for determining what is desired or feasible within a bounded application space such as meeting a specific client's needs. This also often involves taking individual responsibility for being an effective communicator and collaborator. Different perspectives are seen as a "value-added" but not understood in a deep way.

Category 2 (Intentional Learning) distinctly focuses on the process and outcomes of collaborative and situational learning. This category builds off of Category 1 (Working Together) because it represents a process of *improving the conditions needed to work together with people with different training to address complex problems of social and global significance*. As shown in Table 1, a focus on cooperation and collaboration expands to include social learning, a focus on complex problems expands to include social and global elements, and a role evolves from being a collaborator to being a self-directed learner. Awareness of differences changes from recognizing disciplinary differences to respecting the difficulty of disciplinary training and learning at the intersection of differences. Creating opportunities to address complex challenges emerge from intentional learning experiences that involve seeing failure through an opportunistic mindset and having a passion for self-directed learning. Here, personal learning occurs in parallel with efforts to be an educator so that collaborators can co-create a shared understanding.

Category 3 (Strategic Leadership) distinctly focuses on proactively enabling successful cross-disciplinary discovery and innovation through managing and leveraging differences. "Orchestrating" is emphasized explicitly through a self-identified role of being a facilitator at the cross-disciplinary interface and taking the risk of leading projects towards successful outcomes. As shown in Table 1, this is a leadership role that is about enabling the team rather than promoting individual egos. Like Category 1 (Working Together) elements of collaboration and successful outcomes are evident in Category 3; however, an awareness and understanding of what enables success expands to include issues of trust, respect, shared ownership, and inclusivity such that disciplinary, organizational, and cultural perspectives can be synergistic and open up new ways of thinking. Category 3 builds off of Category 2 (Working Together) by *applying prior learning on the challenges and affordances of cross-disciplinary practice to proactively enable successful cross-disciplinary discovery and innovation*. As shown in Table 1, prior experiences have established an awareness of how cross-disciplinary work can break down: differences in perspectives across disciplinary, organizational, and cultural perspectives; poor problem formulation that leads to ineffective and inappropriate solutions; an inability of disciplinary paradigms to meet the needs of complex situations that involve economic and political dimensions; a need to engage a social network of expertise; and the importance of building allegiance, a common vision, and collaborative decision making.

Category 4 (Challenge and Transform Practice) distinctly focuses on questioning practices and boundaries. Where Category 3 (Strategic Leadership) involves leading teams in creating common ground and new ways of thinking, a leadership role for Category 4 expands to include *being a transformative agent as well as being transformed*; where Category 3 focuses on enabling cross-disciplinary discovery and innovation, Category 4 is about *critical reflective practice to enable transformative learning and outcomes (for individuals, teams, stakeholders, and disciplinary practices).* As shown in Table 1, Category 4 involves expanding the idea of "difference" to include lived experiences and recognizing how boundaries between differences are socially constructed. This awareness facilitates critical analysis of the idea of "difference" and a critical exploration into similarities across different perspectives, which leads to new inclusive practices, theories, and identities. The essence of Category 4 is challenging epistemic frames comprised of skills, knowledge, values, identity, and theories of knowledge, as well as honoring differences and how diversity enables transformative thinking and transcending boundaries. Part of this is attuning to the human and contextual aspects of complex problems. This focus on human issues is evident in participatory strategies that engage diverse stakeholders as partners, not just information resources. When human and contextual factors are integrated into the system, the limits of prior views of "good practice" or "good science" are revealed and enable new ways of thinking about system performance. For Category 4 there is a unique and explicit identity of "being cross-disciplinary". This new identity may involve disrespect within and exclusion from prior disciplinary communities as well as seeking out new revolutionary "homes".

## Implications and Future Work

While there are a number of theoretical implications of this study, a central contribution is a developmental framework of cross-disciplinary thinking, acting, and being at the individual level. Overall, the analysis revealed four hierarchically related categories of variation that emphasize (1) the ways "difference", complexity, and goal motivations are experienced and (2) how these experience shape actions and self-perceived identities. Across these variations are themes of epistemological development, respect for “difference”, social networking and participatory acts to create common ground, and identity development (for self and in relation to others). F or example, an expanding awareness of “difference” is associated with epistemological reflection and an evolving respect for different perspectives. This is evident in the ways “difference” is described (from disciplines to organizational structures to cultures to lived experiences), the actions associated with an awareness of difference (as something to be overcome, explored, leveraged and managed, and critically reflected upon and transformative), and how it is characterized in reference to other perspectives (as a relevant “value-added”, as respected as difficult, as respected in decision making, and as honored as a way of knowing). In this way, studies of cross-disciplinary practice may help understand how disciplinary and cross-disciplinary learning interact.

This study also suggests practical implications. In phenomenography, the link between practice and preparation for practice is that the way learners experience a phenomenon *in the past* will form *how they act in the future*. Some of the themes in the results have direct implications for structuring and sequencing learning opportunities. For example, a theme of learning to collaborate as well as learning through collaboration suggests the potential impact of engaging learners early and often in cross-disciplinary collaborations – regardless of their level of disciplinary grounding. Similarly, a theme of intentional learning suggests the potential for connecting self-directed and self-regulated learning strategies or mindsets to a broader agenda of cross-disciplinary learning. At a minimum, the role of variation in developing an awareness of cross-disciplinary practice suggests a critical need to provide learners with diverse experiences to enable an embodied understanding of practice.

## Acknowledgments

The authors would like to thank Shanna Daly, Denny Davis, and David Radcliffe for their insightful contributions, and the National Science Foundation for funding this work (EEP-0748005).

## References

Adams, R.S., Mann, L., Jordan, S., and Daly, S. (2009a). “Exploring the boundaries: language, roles, and structures in cross-disciplinary design teams.” In J. McDonnell and P. Lloyd (eds), *About Designing: Analysing Design Meetings*, 2009, Chapter 16. London: Taylor and Francis Group.

Adams, R.S., Forin, T., Mann, L. and Daly, S (2009b). “Cross-disciplinary practice in engineering contexts.” Proceedings of the *International Conference on Engineering Design* (ICED), Palo Alto, August.

Bowden, J.A. & Green, P. (2005). *Doing Developmental Phenomenography*. Melbourne: RMIT University Press.

Bromme, R. (2000). “Beyond One’s Own Perspective: The Psychology of Cognitive Interdisciplinarity.” In P. Weingart & N. Stehr (Eds), *Practising Interdisciplinary*, pp. 115-133. Toronto: University of Toronto Press.

Daly, S., Mann, L., and Adams, R. (2008). “A new direction for engineering education research: Unique phenomenographic results that impact big picture understandings.” In *Proceedings of the 2008 AaaE Conference*, Yeppoon, Australia.

Dall’Alba, G. (2009). *Learning to be Professionals. Innovation and Change in Professional Development, Volume 4*. Springer Science+Business Media B.V.

Dall'Alba, G. and J. Sandberg (2006). "Unveiling Professional Development: A Critical Review of Stage Models." *Review of Educational Research,* 76(3): 383 - 412.

Galison, P. (1997). *Image and Logic*, 1997. Chicago: The University of Chicago Press.

Klein, J. T. (1996). *Crossing boundaries: Knowledge, disciplinarities, and interdisciplinarities*. Charlottesville, VA: University Press of Virginia.

Klein, J. T. (2004). “Prospects for transdisciplinarity”. *Futures*, 36: 515-526.

Lattuca, L. (2001). *Creating interdisciplinarity: Interdisciplinary research and teaching among college and university faculty*. Nashville: Vanderbilt University Press.

Mann, L. M. W., D. Radcliffe, and Dall’Alba, G. (2007). “Using Phenomenography to Investigate Different Ways of Experiencing Sustainable Design.” In *Proceedings of the ASEE Annual Conference & Exposition*, Hawaii.

Marton, F. and Booth, S. (1997). *Learning and Awareness*. New Jersey: Lawrence Erlbaum Associates.

National Academy of Engineering (NAE) (2004). *Engineering of 2020: Visions of Engineering in the New Century*. Washington DC: National Academy of Engineering, National Academy Press.

Nersessian, N. J. (2006). “The Cognitive-Cultural Systems of the Research Laboratory.” *Organizational Studies*, 27(1): 125-145.

Newstetter, W., E. Kurz-Milcke, et al. (2004). “Agentive Learning in Engineering Research Labs.” Proceedings of the annual Frontiers in Education Conference.

Sfard, A. and A. Prusak (2005). "Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity." *Educational Researcher,* 34(4): 14-22.

# VOLUME 1 AUTHOR INDEX

**D**

**H**

**M**

**R**



**S**

**T**



**V**



**W**