How to study learning processes? Reflection on methods for fine-grain data analysis

Orit Parnafes, Tel-Aviv University, Israel, oritpa@post.tau.ac.il
David Hammer, University of Maryland, College Park, MD, davidham@umd.edu
Loucas Louca, European University, Lefkosia, Cyprus, Louca.L@cytanet.com.cy
Bruce Sherin, Victor Lee, and Moshe Krakowski, Northwestern University, Chicago, IL
Email: bsherin@northwestern.edu, victor@northwestern.edu, m-krakowski@northwestern.edu
Andrea diSessa, University of California, Berkeley, CA, adisessa@soe.berkeley.edu
Daniel Edelson, Northwestern University, Chicago, IL, d-edelson@northwestern.edu

Abstract: This symposium addresses methodological issues in studying children's knowledge and learning processes. The class of methods discussed here looks at processes of learning in fine-grained detail, through which a theoretical framework evolves rather than is merely applied. This class of methodological orientations to studying learning processes diverges from more common ones in several important ways: 1) Attention to diverse features of the learning interaction; 2) conducting a moment-by-moment analysis, zooming in on the fine details of the studied processes; 3) rather than proving or applying a theory, the objective is to make theoretical innovations, or to develop a "humble theory." The challenge of using such techniques is that, by their nature, they do not follow a strongly delineated procedure, especially not the usual sort of coding. This symposium attempts to begin addressing the methodological issues by reflecting on several cases of data analysis.

Introduction

This symposium is framed as a methodological one. Its main objective is to address questions regarding how to study learning processes, with an emphasis on research design that is based on a few in-depth case studies. More particularly, by studying learning processes, we mean, for the most part, studying processes of knowledge development and conceptual change, what usually is referred to as conceptual analysis. We name this class of methods as "knowledge analysis." The class of methods discussed here looks at such processes of learning and knowledge development in fine-grained detail, through which a theoretical framework evolves rather than is merely applied. We summarize here its unique characteristics:

Focus on knowledge and conceptual processes.

Attention to diverse features of the learning interaction: the researcher pays extended attention to the detail of the interaction, without ruling out relevance of any aspects of the studied situation in advance.

Fine grained analysis: the analysis is conducted at a very fine timescale. The researcher does a moment-by-moment analysis of a selection of episodes of interest, zooming in on the fine details of the phenomenon in order to gain insight into these moments of learning. The implication of this fine grain of observation is that this data analysis is usually applied on a small number of cases.

Theoretical orientation: the researcher aims at developing a "humble theory" (diSessa & Cobb, 2004), or extending or modifying an existing theory of learning. The objective is not to prove a theory or apply it, but to make some new theoretical extensions and innovations. Usually, the objective of the research is evolving, and develops from recurring close examination of the data, trying out various theoretical schematizations.

The aim of this class of data analyses is to answer questions of "how and why" with regard to learning. This is a type of research that seeks to uncover mechanisms of learning, to help us gain understanding about how learning comes about and how knowledge develops and changes.

In conducting a quantitative study, e.g., a comparative study, the researcher needs to follow a specific set of steps. The steps may include sophisticated thought in carefully crafting the procedure, and in controlling the variables. However, often, there is only little uncertainty in the method itself. Even in some of the qualitative methods, for example, grounded theory (Strauss & Corbin, 1990) prescribed guidelines are given. The class of methods we deal with does not follow a strongly delineated procedure, and the range of inventiveness is large. Schoenfeld (1992) describes such occasions as a situation in which one focuses on events that one wishes to understand, but for which extant methods are not satisfactory. In that case, one should invent a new method for data analysis suited for the question at hand. This process is usually accompanied by the evolution of a new "humble theory" (diSessa & Cobb, 2004).

What constitutes rigorous analysis in these contexts, when there is no standardization of approach? This symposium attempts to begin addressing this question. We do not intend to provide a systematic set of guidelines for such a method. Rather, we will share some reflective notes about processes of data analysis – a reflection that aims to uncover some of the tools of the trade of this kind of research.

Hammer and Louca begin the session with a critical examination of coding practices in qualitative research. Their argument takes a stand against the practice of coding data, and directly treating the codes as data in subsequent analyses. Since each code is itself a small claim about data, not data itself, then the justification and strength of those claims can vary greatly. Their presentation sets the ground for alternative forms of analysis, exemplified in the other contributions.

Sherin, Lee, and Krakowski discuss the development of a framework concerning how to interpret interview data designed to get at students' science conceptions. The idea of the framework is that in order to look at an interview and figure out what knowledge a student has, one has to understand the interview itself as a dynamic interaction, during which a student assembles ideas. The process through which their theoretical framework evolved will be analyzed.

DiSessa will schematize the process of developing a particular theoretical construct. He will first schematize the general process of developing a theory of knowledge out of data, and then address specifically the issue of how to "observe" knowledge, after a sketch of a theory exists. The analysis includes a classification of principles for identifying knowledge elements.

Finally, Parnafes will reflect on a process in which a theoretical framework is chosen as a starting point for a specific context and the way it gets appropriated to produce insights in the context of the research. This is a slightly different process, involving distinct challenges, unlike the case where the theoretical framework is original. She will exemplify the process in which theoretical innovations evolve through the negotiation between a selected "flexible theory" and data.

The participants of the session would address some of the following issues:
Issues of theoretical framework
Producing rich and interesting data
Approaching raw data – processes of selection and focus.
Data interpretation in conjunction with the theoretical framework.

Summaries of the particular contributions

Challenging accepted practice of coding

David Hammer and Loucas Louca

It has become standard, accepted practice in education research to collect qualitative data and "code" it by a scheme developed within or prior to the study. Within this practice, it is important to establish that the scheme satisfies a threshold of inter-rater reliability, typically 80%, and most studies involve the additional step of multiple coders resolving disagreements through discussion. The outcomes of these analyses are then treated as data for further analysis. In this way, researchers treat coding as a process of data reduction, from the raw data contained in video records, transcripts, or other artifacts, to the numerical counts by the categories of the framework. Those numerical counts become the data for statistical calculations in support of various claims.

Recent research on student argumentation in science provides examples, in work across three projects as published in five different journals (Erduran, Simon, & Osborne, 2004; Felton & Kuhn, 2001; Kuhn & Udell, 2003; McNeill, Lizotte, Krajcik, & Marx, 2006; Osborne, Erduran, & Simon, 2004). In each case, the researchers develop a scheme for coding qualitative data (student discourse drawn from classroom or clinical settings), reported inter-rater reliability as exceeding 80%, and used the results of the coding schemes as data for statistical analysis. The articles that report the research present only brief samples of the qualitative data to illustrate and warrant the coding schemes. There is no question that these projects reflect valuable work or that there are insights to be gained from their analyses. We argue, however, that the practice of treating the numerical results of coding as data is fundamentally flawed, for several reasons.

Codes are claims about data

To begin, coding is essentially a process of making claims about data, and the quality of the support for those claims varies considerably. That the quality of the support varies is apparent in the rates of agreement among independent raters: That multiple coders, even after training, may disagree in up to 20% of the coding decisions is evidence that in many instances the coding decisions are difficult to make — likely including instances when the coders agree. The difficulties of those decisions may result from a mismatch between the categories of the scheme and the data — given the choices of the framework, coders may not find any of them a clear fit, even if they agree over which is the best. Moreover, inter-rater reliability is almost always assessed among researchers collaborating on the project, who may share tacit understandings or biases that lead them to agree with each other on matters that outsiders would find problematic. In all of these respects, the quality of the support for coding decisions varies, in ways that are obscured when research presentations do not provide significant access to the data.

The problem is that many studies such as those cited above, present the results of coding as if it is data itself, including in calculations of effect sizes and statistical significant, treating the numerical counts in coding categories as one might treat numerical counts of student selections on multiple choice exams. In the latter case, it is the students making the selections; in the case of coding qualitative data it is the researchers. By treating the results of coding as data, researchers leave the process of their own decision making out of their calculations. The fact that inter-rater disagreement can be as high as 20% makes clear that the process of coding introduces a new source of variance, but in many studies, including those cited above, this variance is not included in statistical calculations. In other words, the calculations do nothing to differentiate codes that were produced by discussion to resolve disagreement, or those that represented difficult decisions, from those in which the choices were unambiguous.

For these reasons, the process of moving from qualitative data to quantitative counts in the categories of the scheme should not be understood or treated as data reduction. Research articles that present empirical findings must present sufficient data to support their claims, and codes are not data. We recognize that this criticism might, in principle, be leveled against transcripts or even video recordings, as the transcribers and videographers invariably used judgment to make decisions. Still, transcripts and video recordings are without question closer to the phenomena of interest than are the results of coding.

Coding schemes both guide analysis and limit discovery

We do not argue against the development and use of coding schemes. They can and do play an important role in guiding analysis. To develop a coding scheme is to formulate articulate description of categories, with inter-rater reliability a means of assessing the clarity and sufficiency of that articulation. It is a process of synthesis, with the coding scheme representing a useful target epistemic form (Collins & Ferguson, 1993), useful in focusing and mediating researchers' interactions with data and with each other. It is an achievement to arrive at a scheme that multiple researchers can use to arrive at similar conclusions, and in that respect the scheme itself is a finding (Marton, 1988). One implication is that it is especially useful for researchers to explicate the development of the scheme; Erduran, Simon & Osborne (2004), for example, provide an account of their adaptation of Toulmin's framework to the purpose of analyzing data from science classrooms.

Moreover, the achievement of a scheme can be of value to others, as the application of a coding framework to data presses researchers to examine it for particular features. Research on argumentation, for example, has clearly been advanced in the projects cited above, as the respective frameworks provide; in other work (Louca & Hammer, in preparation), we analyzed an episode of children's argumentation in part by applying the coding schemes cited above, and we showed how these schemes provided insight into the data. We discuss how the application of coding schemes steered us to notice particular student utterances, how it led us to examine the data for evidence of meanings we would not otherwise have considered.

The accepted practice of treating codes as data, however, often involves a premature reification of coding schemes. There is value in the iterative development, application, and refinement of a coding scheme, but practices in education research move too quickly to apply these schemes as tools to do other work. For example, having developed a coding scheme, researchers may utilize it as part of a different iteration, one of development, implementation, and refinement of curriculum materials. In each of the three projects cited above on argumentation, the researchers proceed to apply their coding scheme to support claims comparing the effectiveness of different pedagogical interventions for students. It is this use we argue is inappropriate.

How to conceptualize rigor in data analysis?

The examples we picked of research that treats coding results as data are certainly not unusual; we picked them simply because they connect to our interest in children's argumentation. The practice we are criticizing is in wide use across education research, largely because it provides a sense of methodological rigor and structure. If as we argue the results of coding cannot be understood as data, then this sense is misleading, and researchers need to look to other means of pursuing and assessing rigor in the analysis of qualitative data, without recourse to problematic quantification.

Coding schemes can be a part of that rigor, for the reasons we mentioned above: Their development and application focus and mediate researchers' interactions with each other and the data. We will discuss our use of the coding schemes from the literature, giving examples of the dynamic they supported in our analysis such as to help us discover and articulate claims we could support and claims we could not. Coding schemes can support a similar dynamic on a larger scale in the research community, but we need to build practices around a substantive exchange of *data*.

Complexity and compromise: Treating interviews as dynamic interactions

Bruce Sherin, Victor Lee, and Moshe Krakowski

For greater than two decades, researchers in science education have been concerned with studying students' *intuitive science knowledge*, the informally-gained knowledge of the natural world that students bring to the study of science. Although intuitive science knowledge has been the focus of extensive research efforts, across many domains, and employing many techniques, some of the most fundamental questions continue to be hotly debated. In many cases, these debates seem to hinge on issues of methodology. This can be illustrated with a prominent example.

In 1992, Vosniadou and Brewer published a seminal article in which they described children's models of the shape of the Earth, and discussed how these models seem to change as students age (Vosniadou & Brewer, 1992). Their results showed that each individual student seems to possess one of a small number of mental models of the Earth's shape, and that they apply these models consistently. Recently, however, a collection of articles, describing differing results, have called Vosniadou and Brewer's claims into question. In part, the differences seem to stem from the use of differing methodologies. Vosniadou and Brewer asked children relatively open ended questions. It seems that children give different answers when they are given concrete models to manipulate (Siegal, Butterworth, & Newcombe, 2004), or if the questions asked required children to selected from a fixed list of candidate models (Nobes et al., 2003; Vosniadou, Skopeliti, & Ikospentaki, 2004).

Arguments surrounding this debate seem to center on determining which of these diverse interviewing methodologies provides a more "true" picture of children's underlying knowledge. However, we believe that this is not the most appropriate stance to adopt. No type of interviewing technique provides a transparent window into the thinking of students. Instead, we believe, the interviews themselves must be analyzed as dynamic processes during which the child interacts with the interviewer, and constructs responses drawing on their own knowledge, and prompts from the interviewer.

If one is serious about adopting this new analytic stance, there are a number of serious challenges. When students are interviewed about natural phenomena, they can often draw on a huge body of relevant knowledge. Students know, for example, that the Earth appears flat, that it looks round in pictures from space, that Australia is far away, that you can dig to China, etc. Furthermore, the selection of knowledge a student draws upon at a given moment might depend in a very sensitive manner on exactly what questions the interviewer asked and how the interview happens to unfold.

Modeling this complex process requires that we make a number of compromises. We must be willing to engage in guesswork, and to tolerate a significant amount of tentativeness and unreliability at the local level of our analysis. Furthermore, in order to somewhat rein in this complexity, our theoretical frameworks must be kept "humble." Our particular "humble" framework employs two simple and neutral theoretical constructs to describe intuitive knowledge. We use the term *node* to refer to the elements of knowledge that appear in our analysis. Though we use a single term, we understand nodes to be of diverse types and to live at multiple levels of abstraction. Our second theoretical construct is what we call a *mode.* A mode is a recurrent pattern in the activation of nodes.

In order to illustrate these challenges and compromises, and our own approach for negotiating them, we will describe our analysis of interviews in which students were asked to explain the seasons. In doing so, we will draw on a corpus of approximately 40 interviews with middle school students on this topic. These interviews indeed seem to involve complex interactions, and students draw on a substantial body of knowledge. Our most recent analyses involve an attempt to code the interviews in terms of a set of 109 distinct nodes. In part because the number of codes is so large, it is extremely difficult to code the interviews in a reliable manner, and a great deal of guess work and inference is required. We will demonstrate these challenges. But we also hope to show that it is possible to draw important conclusions with a great deal of confidence; indeed, we believe the level of confidence can be much higher than with more "reliable" coding techniques.

How to Observe Knowledge

Andrea diSessa

This work aims contribute to qualitative methods of data analysis. However, it is also committed to the idea that it is difficult or impossible to "observe" things such as knowledge without an adequate view of the *nature* of those things. [In jointly developing theory and empirical results, this work is comparable to "grounded theory," (Glaser & Strauss, 1967).] Given the state-of-the-art concerning models or theories of knowledge – especially certain kinds of knowledge, such as intuitive knowledge – this means practically that we must simultaneously develop theory and empirical results. I use the term *knowledge analysis* to describe joint inquiry on (1) the nature of certain kinds of knowledge (theory development, concerning *form*) and (2) whatever particular things people know of that kind (empirical work, concerning *content*).

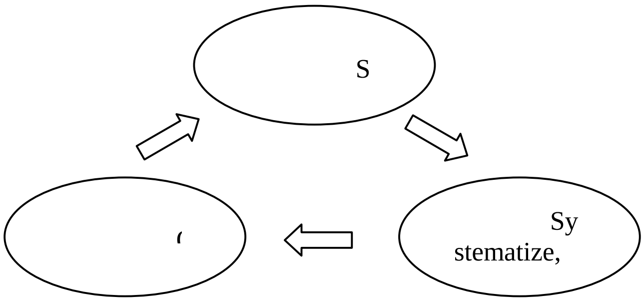

Figure 1. The process of jointly developing theory and empirical observations.

I begin with an overview of the process of jointly developing both theoretical and empirical results concerning knowledge. Figure 1 shows a cyclical process of observing intellectual performance, schematizing that performance in terms of the knowledge that it betrays, and then systematizing and formalizing the characterizations of knowledge that have been produced. The process is highly bootstrapped, and in practice none of the phases can get very far without the others. Cycles may be small or large, and, while the sequence from observation, to schematizing, to putting all the pieces into a systematic theory or model may seem natural, in practice any phase may follow any other, and for a variety of reasons.

Each phase is, by itself, complex. *Observation*, for example, entails the creative process of inventing and instantiating conditions under which important things can be "observed." *Schematizing* typically involves an iterative loop of describing, looking for data supporting for the validity of that description, looking for undermining data, and improving the description iteratively. The third stage, the most obvious locus of theory development, depends on many things, prominently the general form of theory that researchers are trying to produce.

I focus here on a small piece of this general process, which is how data are interpreted to "observe" knowledge after significant bootstrapping has taken place and after a good sketch of the theory exists. The model of observing presented here is abstracted from, and will be exemplified by, an extensive study of intuitive knowledge, diSessa (1993). In its briefest form, this study hypothesized that intuitive knowledge in physics consists of a large number of relatively independent schemata, called p-prims. P-prims are hypothesized to develop to explain and predict the outcomes of commonly experienced events, for example, that "working harder gets a greater outcome." They are activated implicitly in students' thinking about a situation, are not easily expressible in language, and come to serve important roles in learning technically correct physics.

The model of "observing" used for this work had to accommodate substantial difficulties in identifying knowledge elements like p-prims, which are transient, cannot be articulated directly in language, and are likely to involve sensitivity to categories such as "agency" in inanimate interactions (Talmy, 1988), which are unconscious and often expressed in "bodily" terms (such as perceived, personal effort). In short, one produces an argument for one's interpretation that is judged according to a set of explicit principles. In detail, *the nature of each element identified should be substantiated through an extended argument involving marshalling data from multiple occasions of use and non-use, and judged according to a set of "principles of good form."* DiSessa (1993) details about 20 such principles, of which a few are listed below. The principles, as a whole, bring a very wide range of considerations (including their developmental trajectories and fine details of their dynamic use) to bear on triangulating the nature of particular knowledge elements.

General Principles

Invariance – A description of a p-prim should explain both when it is used and when it is not used.

Redescription – In the process of describing a p-prim, one should generally consider many alternatives and produce a "competitive argument" that the proposed description is best.

Somewhat Specific Principles

Coverage – The full set of p-prims should be expected to cover a very wide range of, if not all, of the physical phenomena that are encountered in everyday experience (as opposed to a small class of specialized problems, which is the typical locus of developmental and "misconceptions" research).

Principle of the body – P-prims are likely to use the "vocabulary" of bodily experience in describing the world.

Specific Principles

Obviousness – Everyday happenings that are taken to be obvious need to be explained by the invocation of particular p-prims.

Impenetrability – P-prims are self-explanatory, and subjects can almost never produce justified arguments for their validity.

After sketching this program, the presentation will include an "evaluation" of the successes and failures of this methodology, including speculation about the roots of its good and bad characteristics. For example:

Positive Characteristics

The program supported at least one extended and fairly successful inquiry.

It also has been taken up by a few other researchers (e.g., Sherin, 1996; Azevedo, 2005).

Negative Characteristics

The long-term development of the program—for example, in community argument that develops a better and more refined common set of principles—has been minimal.

These principles, or even the whole program, may be too specific to the kind of knowledge on which it was initially used to spread widely.

Producing an extended argument for each proposed knowledge element may be, in practice, unwieldy as a form for scientific social interchange.

Theoretical framework – processes of appropriation and evolution

Orit Parnafes

In this part, I will reflect on a process in which a theoretical framework is chosen for a specific analytical context and the way it gets appropriated, extended and changed to produce insights in the context of the research. This is a slightly different process, involving distinct challenges, unlike the case where the theoretical framework is original and evolves "from scratch" out of the data.

The research on which this reflection is done concerns the development of students' conceptual understanding of a class of physical phenomena, natural harmonic oscillation, through the mediation of dynamic and interactive representations (computer-based simulations). The goal of this research is to construct a model that describes mechanisms of developing conceptual understanding through the mediation of external representations. The process of students' developing understanding through the use of representations is analyzed from a "knowledge in pieces" perspective (diSessa, 1983; 1993), using a developing theory called *coordination class theory* (diSessa & Sherin, 1998).

The data comes from observations of pairs of middle and high school students engaging in explorations of the phenomena of natural harmonic oscillation. The students first explore the physical phenomenon using physical devices and then explore the same phenomenon with multiple computer-based representations. Two questions guided the investigation: 1. how would conceptual change be "seen" when it happens in this context? 2. how do the changes observed relate to the use of the representations?

This is an attempt to schematize the process of appropriating a theoretical framework to a specific analytical context. The schematization of the process is combined from several phases: the incubation phase, the theory-data negotiation phase, and the theory appropriation phase.

The incubation phase

This phase is characterized by massive examination of the data, both video and transcripts, over and over again. During this phase a substantial number of ideas and insights are drawn to produce a list of proposals and issues for consideration for subsequence analysis. An example of an issue that emerged, among many other issues, followed the recognition of a number of situations characterized as "breakdowns". The detection of breakdowns put into question how central might breakdowns are in analyzing the development of conceptual change through the use of representations. Several questions were raised with respect to breakdowns: what are the different processes that lead to breakdowns? What is the role of the representations in generating such breakdowns? And, what are the different ways students deal with breakdowns?

This phase is unsystematic and exploratory in nature. However, it is accompanied by local patches of systematic data analysis to figure out notable patterns. For example, recognizing episodes of breakdowns, several segments of data were chosen to see the circumstances and process that led to these breakdowns. Those are preliminary attempts to intimately getting to know the data, explore different issues and getting a feel for various patterns. These initial explorations are the seeds for subsequent theoretical constructs, as they become more precise and systematic.

The theory-data negotiation

The theoretical framework evolved through two simultaneous processes. The first process is a continuation of the exploratory phase described before – a process in which ideas, insights and questions evolve from the data, in a grounded theory fashion—that is, always stimulated by and accompanied by proposed examples. The second process is the examination of existing theories that link to issues and questions of concern. Their usefulness is checked on segments of data to see what insights they bring. The selection of a specific theory for applying on a body of data to answer specific questions, should take into account the following:

Fit with the researcher's theoretical commitments and perspectives.

The theory should include parts or features that could address some of the main questions of the research, or at least be insightful in some way. The theory may appear to match the kind of questions asked in the research, but whether it really is insightful and meaningfully addressing the questions often cannot be judged until one tests the application of the theory on the data. Thus, the process may be long and iterative.

After examining several theories, I selected the "coordination class" theory (diSessa & Sherin, 1998), as a theoretical framework for my research. Even though coordination class theory already existed, it wasn't used as a grand theory in the sense that grounded theory describes (Glaser & Strauss, 1967). Rather, it was used as a theory in development, where parts of the theory were elaborated and evolved from the present data. Once a certain theory is selected for use, the development of the framework is turned into a negotiation between a "flexible theory" and the data, with its insights and ideas that were continually collected. How do these ideas relate to the constructs of the theory? How potentially insightful could the theory be with regards to the issues raised? The application of the new framework on a range of segments of data and the examination of the ways it support the data interpretation and how insightful it is, brought up some refinements and adjustments as described in the next section.

Appropriating and developing the theoretical framework

Theory appropriation is a process that takes place through continuing application of aspects of the theory on segments of the data, while various issues and insights from the exploratory process feed in and are taken into account in appropriating and refining the theory. In this research, the theory was refined in the following ways: (1) existing theoretical constructs were refined to match the particular context under exploration; (2) Undeveloped constructs were developed through their application on the data; (3) relevant parts of the theory were used in innovative ways. It is important to note that coordination class theory invites extensions and changes of this sort, because it is intended deliberately as a sketch and has an explicitly tentative nature.

A major innovation was generating a model that describes fine-grain mechanisms that drive the process of developing understanding through the use of representations. The model evolved through the negotiation between the existing theory and the various issues and insight evolved from the open exploration of the data. On the one hand, the mechanisms grew out of central issues that were raised in the exploratory process. For example, breakdowns serve an important role in the model. One of the mechanisms (the "challenging mechanism") is based on the idea of breakdowns in a "softer" version. On the other hand, the mechanisms also tightly related to theoretical constructs of coordination class theory. The evolution of the mechanisms through this negotiation will be described and exemplified in detail.

Discussant

Daniel Edelson, Northwestern University

References

Azevedo, F. (2005). Serious play: A comparative study of learning and engagement in hobby practices. Unpublished PhD dissertation. UC Berkeley.

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist, 28*(1), 25-42.

diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 15–33). Hillsdale, NJ: Erlbaum.

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., & 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. *Journal of the Learning Sciences,* 13(1), 77-103.

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.
Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes, 32*(2-3), 135-153.
Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine.
Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*(5), 1245-1260.
Marton, F. (1988). Phenomenography – A research approach to investigating different understandings of reality. In R. R. Sherman & R. B. Webb (Eds.), *Qualitative Research in Education: Focus and Methods* (Vol. 21, pp. 143-161). Londer: Falmer.
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.
Nobes, G., Moore, D. G., Martin, A. E., Clifford, B. R., Butterworth, G., Panagiotaki, G., et al. (2003). Children's understanding of the earth in a multicultural community: mental models or fragments of knowledge? *Developmental Science, 6*(1), 72-85.
Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994-1020.
Schoenfeld, A.H. (1992). On paradigms and methods: What do you do when the ones you know don't do what you want them to? *Journal of the Learning Sciences, 2*(2), 179-214.
Sherin, B. (1996). The symbolic basis of physical intuition: A study of two symbol systems in physics instruction. Unpublished PhD dissertation. UC Berkeley.
Siegal, M., Butterworth, G., & Newcombe, P. A. (2004). Culture and children's cosmology. *Developmental Science, 7*(3), 308-324
Strauss, A.L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.
Talmy, L. (1988). Force dynamics in language. *Cognitive Science*, *12*, 49-100.
Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*(4), 535-585.
Vosniadou, S., Skopeliti, I., & Ikospentaki, K. (2004). Modes of knowing and ways of reasoning in elementary astronomy. *Cognitive Development, 19*(2), 203-222.