Open community authoring of worked example problems

Turadg Aleahmad, Vincent Aleven, Robert Kraut, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA
Email: turadg@cmu.edu, vincent.aleven@cs.cmu.edu, robert.kraut@cmu.edu

Abstract: Open collaborative authoring systems such as Wikipedia are growing in use and impact. This research examines how to create a collaborative authoring community for educational resources. We describe and evaluate a novel tool for community authoring of worked examples, in the task of making instruction for a specific math skill. Participants were professional teachers (math and non-math) and amateurs. We find that while there are differences by teaching status, all three groups made contributions of worth.

Introduction

The production of instructional material typically involves a closed system; one person or cohesive group produces each artifact, be it a paper textbook or computer tutor. But over the last decade open development models have emerged. Open source software development is a vibrant enterprise and has produced the Linux operating system, the Apache web server and many other commercially important products. Wikipedia brings a similar model to the organization of knowledge. While Wikipedia has benefits for education, it is designed to teach "what" and not so much "how". To develop open educational resources may require new models of collaboration. The experiment described here is part of a larger program of research through design of a web-based community to support a full life cycle for open educational resources: 1) Generation in which volunteers submit material, 2) Evaluation in which members of a community rate and critique generated content, 3) Use by teachers and others through customized instructions, worksheets and handouts, and 4) Improvement in which volunteers makes the contribution better. The vision is a model of development that is cheaper than existing methods, leads people to think more about learning, and produces the best resources. This study is solely on the Generation phase of the cycle.

In this experiment we explored how to generate educational content through an open source model. We applied the model to the generation of worked example problems. A worked example problem consists of a problem statement and a sequence of steps detailing the solution and how to reach it. They are in many cases more effective than exercises in helping students learn. Worked-examples both instruct and help to foster self-explanation. (Renkl and Atkinson, 2002) They also complement existing practices and systems, such as interactive tutoring (McLaren et al., 2006; Schwonke et al., 2007). They are also versatile in that a worked example can be presented to a student as a full problem and solution, as a problem with a partial solution, or as simply a problem, as on a homework sheet. In this experiment, we asked participants to develop worked examples for understanding and applying the Pythagorean Theorem. The system was made open to all comers and we compared the contributions of professional teachers (math and non-math) and amateurs.

Methods

The experiment took place over the internet through interactions with a prototype web application. Participants were presented a page explaining what a worked example problem is and that their task is to create a worked example problem to teach the Pythagorean Theorem. They were provided help in the task through three simplified pedagogical principles and a search box to look up information in Wikipedia.

The next screen was the authoring tool. Tabs at the top guided the user in the task. Authors wrote the text of the problem statement in a large box to the left and drew their diagrams using a simple pen tool in a box to the right. After developing their problem statement and diagram, they clicked "Add Step" to add each step of the solution. The solution area had three columns, one for the work for the student to perform towards the solution, one for the explanation of that work, and one for an illustration.

The URL to participate was advertised on various web sites both related to education and not. During the experiment 1427 people registered on the site to participate. After seeing the task in detail most did not continue, but 570 participants did use the system to submit 1130 contributions. Participants were asked at the outset to select their teacher status: math teacher, other teacher, not teacher.

The submissions were filtered automatically on simple quantitative criteria, as a real community site would. The base criteria were: Problem length between 50 and 1000 characters, and at least one solution step provided. 551 submissions from 281 participants met these criteria.

These machine-vetted submissions were later coded by two math teachers. They rated the worth of the problem statement, the solution work and the solution explanation on a 4-point Likert scale, with ratings labeled: Useless, Easy fix, Worthy or Excellent. Both were blind to the source of the submissions. Each took a median time of 36 seconds per submission. The expert codings were converted to quantitative values: 0=Useless, 1=Easy fix, 2=Worthy, 3=Excellent. The ratings of the two coders on the problem statement were combined to a “statement quality” scale with an inter-rater reliability alpha of 0.61. The ratings of the two coders on the solution work and explanation were combined into a “solution quality” scale with an alpha of 0.79.

Results

We looked at the quality of all problems submitted and the work needed to classify them. Of 1130 raw submissions, 11% of whole problems (statements with solutions) were classified as Worthy, meaning that they are fit for use immediately. 39% were at least Fixable, meaning that they would be valuable with some additional effort. In general the statements were of higher quality than the solutions. 27% of statements were Worthy and 4% were Excellent as is. Classification into the Filtered category was a trivial computation. Classifying the three contribution components into the four-rating rubric took experts a median time of 36 seconds per contribution.

We looked at the quality of each contribution as a whole, revealing no superior quality by teacher status ($F(2)=1.53$, $p=0.22$). Further analysis revealed that the effect on quality of teacher status interacted with the problem component. Math teachers were best at writing problems statements. A comparison across teacher status showed a marginally significant effect ($F(2)=2.39$, $p=0.093$). Math teachers’ contributions rated at $M=1.79$, followed by amateurs ($M=1.45$) and other teachers ($M=1.45$). A comparison of math teachers with the rest showed a significant effect ($F(1)=4.80$, $p=0.015$, one-tailed). Surprisingly, amateurs were best at writing solutions. A comparison across teacher status showed a marginally significant effect ($F(2)=2.73$, $p=0.067$). Amateurs did best ($M=0.72$) followed by math teachers ($M=0.60$) and then other teachers ($M=0.48$). A comparison of amateurs with the rest showed a significant effect ($F(1)=4.87$, $p=.028$).

Discussion

Through automated methods and software supports for human judges, all the contributions were rated easily. In a short amount of time about 1500 people registered to contribute to a commons of educational materials. While not all came with the same intentions, over half walked away after determining that they did not want to participate fully. Of the raw submissions made, over half were trivial to filter by simple automated methods. Of the remaining, a novice and a veteran teacher were able to rate each of them on three attributes in less than a minute each. About 1/10th were ready to help students learn without needing any modification. Many more were rated as fixable, meaning that with some additional work they would be ready. Statements were the highest quality components and solutions were the most difficult parts to author well.

Teacher status had an important impact on the quality of the components of contributions. Math teachers were best at authoring problem statements. Surprisingly, amateurs authored the best worked solutions. Perhaps this is because they are better able to adopt a student’s perspective. Math teachers performed worse than amateurs but better than non-math teachers. Perhaps this is because their pedagogical content knowledge helps compensate (but not fully) for their expert blind spots.

Overall, it is clear that, at least for worked examples of the Pythagorean Theorem, participants of all teaching statuses were likely to make contributions of value. Math teachers do a better job at some parts of the process, but even laymen do fairly well. This is fortunate because there are many more amateurs in the world than math teachers. In this study each participant made each contribution independently, but the best resources may come from collaboration. For example, a math teacher writes a problem statement and an amateur writes the solution. Educational content systems can benefit from opening the channels of contribution to all comers.

References

McLaren, B. M., Lim, S., Gagnon, F., Yaron, D., & Koedinger, K. R. (2006). Studying the effects of personalized language and worked examples in the context of a web-based intelligent tutor.

Renkl, A., & Atkinson, R. K. (2002). Learning From Examples: Fostering Self-Explanations in Computer-Based Learning Environments. *Interactive Learning Environments*, *10*(2), 105-119.

Schwonke, R., Wittwer, J., Aleven, V., Salden, R., Krieg, C., & Renkl, A. (2007). Can tutored problem solving benefit from faded worked-out examples. European Cognitive Science Conference, 23–27.

Acknowledgments

This work was supported in part by Graduate Training Grant awarded to Carnegie Mellon University by the Department of Education (# R305B040063). The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through "Effective Mathematics Education Research" program grant #R305K03140 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.