Are Worked Examples an Effective Feedback Mechanism During Problem Solving?

Leena Razzaq, Neil T. Heffernan, Prawal Shrestha, Xing Wei, Ashish Maharjan and Cristina Heffernan
(leenar@cs.wpi.edu, nth@cs.wpi.edu, prawal@wpi.edu, wx6872@wpi.edu, maharjan@wpi.edu, ch@wpi.edu)
Worcester Polytechnic Institute, Department of Computer Science
100 Institute Road
Worcester, MA 01609 USA

Abstract

Current research in learning technologies has found both interactive *tutored problem solving* and presenting *worked examples* to be effective in helping students learn math and science. However, which information presentation method is more effective is still being debated among the cognitive science and intelligent tutoring societies and there is no widely accepted answer. This study compares the relative effectiveness between these two strategies when they are used as a feedback mechanism. Controlling for the number of problems, we presented both strategies to groups of students in local middle schools and the results showed significant learning in both conditions. In addition, our results are more in favor of the tutored problem solving condition as it showed significantly higher learning. We propose that the level of interactivity plays a role in which strategies are more effective.

Keywords: Tutored problem solving; Worked examples; Intelligent tutoring systems; Cognitive load.

Introduction

Students are often taught new material in mathematics by first being introduced to the principles needed to understand the new material, then worked examples that show how to use the principles to solve related problems and finally, practice problems for the students to work on. Traditionally, teachers often present only a few examples and assign a large number of practice problems. Likewise, mathematics learning technologies often focus heavily on tutoring step-by-step problem solving with positive learning results (i.e. Cognitive Tutors [1], Andes [18], Mastering Physics [20], and the ASSISTment System [13]) rather than presenting information about principles or presenting many worked examples.

Cognitive scientists have been interested in the role of worked examples in reducing cognitive load and helping students to learn, and there have been numerous studies on the effectiveness of worked examples [8, 14, 15, 16, 17]. Sweller and Cooper [17] presented evidence that supported their hypothesis that worked examples helped novices to acquire "schemas" which they defined as "mental constructs that allow patterns or configurations to be recognized as belonging to a previously learned category and which specify what moves are appropriate for that category." It appears that novices who have not learned the required schemas have to depend on superficial search strategies in solving problems [10] while experts can choose the next appropriate step based on their ability to correctly categorize the problem.

Sweller and Cooper's work suggested that (untutored) problem-solving practice did not help students to acquire schemas as efficiently as the use of worked examples perhaps because of the change of focus from "goal-directed problem-solving" to "problem-state configurations." Kalyuga, Ayres, Chandler & Sweller [8] presented results that point to a benefit of using worked examples early on and then using problem-solving practice later as students show more understanding. Likewise, Schwonke, Wittwer, Aleven, Salden, Krieg & Renkel [16] found that students learned more from gradually fading worked examples to tutored problem solving than from tutored problem-solving alone. In Schwonke et al.'s work, the fading of worked examples was the same for all students and did not depend on their demonstration of understanding.

More recently, Salden, Aleven, Renkl and Schwonke [15] experimented with an adaptive fading scheme where worked examples were gradually faded to tutored problem solving when students showed understanding based on their self-explanations. Salden et al. found evidence that adaptively fading worked examples was more effective than fixed fading.

This study investigates whether students in a classroom setting will benefit more from interactive tutored problem solving than from worked examples given as a feedback mechanism. We also attempt to determine whether students will differ by ability. We expect that more proficient students will benefit more from worked examples than less proficient students. We used the ASSISTment System, described in the next section, to test our hypothesis.

The ASSISTment System

The ASSISTment System [13] offers instruction to students while providing a more detailed evaluation of their abilities to teachers. The ASSISTment System is able to identify the difficulties individual students are having as well as for the class as a whole. Teachers are able to use this detailed data on their students to tailor their instruction to focus on the areas that students are struggling with as identified by the system. Unlike other assessment systems, the ASSISTment system also provides students with tutoring assistance while the assessment information is being collected.

An ASSISTment is the basic unit of our tutoring application. It consists of a single main question which students are asked to answer. Most ASSISTments are based on problems taken from the Massachusetts Comprehensive Assessment System (MCAS) tests, which is the Massachusetts state test that students take every year from grades 3 – 10. The system is primarily used by middle- and high-school teachers throughout Massachusetts who are preparing students for the MCAS tests.

Currently, there are over 3000 students and 50 teachers using the ASSISTment System as part of their regular math classes. We have had over 30 teachers use the system to create content.

Experimental Design

This experiment was conducted with 8th grade students in three local middle schools located in central Massachusetts. One of the schools was suburban, while the other two were urban. Over 80% of the students who participated were from a school which according to its state test scores is in the bottom 5% in the state and has been labeled by the No Child Left Behind Act as not making adequate yearly progress. The experiment took place in the months of April and May of 2008 at the computer labs of the respective schools. The students who participated in this experiment were exposed to both conditions: tutored problem solving and worked examples. They were given problem sets to work on and their actions were logged which was later analyzed.

For the experiment we created nine problem sets, each consisting of four to five ASSISTments. All of the main questions of the ASSISTments were taken from 6th Grade MCAS tests for Mathematics (2001 – 2007) focusing on the Patterns, Relations and Algebra section, which concentrates on different mathematics skills: populating a table from a relation, finding a missing value in a table, using fact families, determining equations for relations, substituting values into variables, interpreting relations from number patterns and finding values from a graph.

Each problem set in this study was a collection of ASSISTments grouped into three sections: pre-test, experiment and post-test. For the experiment, students were considered to have completed a problem set only if they finished every part of it. We used the gain score from pre-test to post-test to determine whether students had learned anything from the conditions.

When students start a problem set, they are first given a pre-test problem. The pre-test is an ASSISTment with a single question, and does not include any form of help or hint.

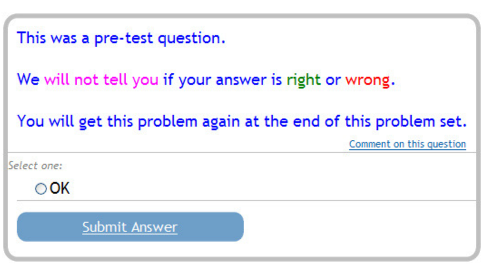

Figure 1: Message receive after students answer pre-test.

In order to make sure that the students understand what is happening, we inform them that the question was a pre-test and that they will not receive feedback on whether their answer is right or wrong. They are also informed that the question will be repeated at the end of the problem set. Figure 1 shows the screenshot of the response that students receive after they answer the pre-test.

In the pretest, students are allowed only one attempt to answer the question, so the first answer they provide is considered as the final answer for the pre-test and it cannot be changed. After the one question pre-test, students are presented with the first question from a randomly chosen condition. The computer randomly assigns either the tutored problem solving or the worked example condition to the students. This part consists of two or three ASSISTments all having the same condition. Within the two conditions, students do the same number of questions, so the content of the questions were held constant between the conditions. Finally, when the students finished all of the ASSISTments in the experiment section, they are given the post-test which is the pre-test question repeated. Also like the pre-test, the first response of the student is recorded and used for analysis. However, unlike the pre-test, we do inform the students regarding the correctness of their answer. Learning can be assessed by comparing the results of the pre-test and the post-test.

In the tutored problem solving condition, students who get a problem wrong are asked to answer a set of questions that break down the main problem into steps. Students can ask for hints on each step if they need more help. For example, in the problem shown in Figure 2 on the left, the student is asked to find the next number in a sequence. If the student provides a wrong answer or presses on the "Break this problem into steps" button, he/she will be directed to the first scaffolding question, which helps the student to understand the first step to solve the original problem. If the student presses on the "Show me a hint" button, hints will be shown one by one until the student reaches a "bottom-out hint" or the answer to this scaffolding question. After this, the student is directed to the next scaffolding question. The number of scaffolding questions depends on the complexity of the original question. At the end, the student is expected to understand how to do the original problem step by step.

In the worked example condition, when a student gives an incorrect answer or presses on the "Break this problem into steps" button, a problem that is similar to the main question is shown solved step by step. As such, the students will have a pattern to follow in order to solve the problem. The worked example condition is shown in Figure 2 on the right. The student is asked to read through the worked example and choose "I have read the example and now I am ready to try again" when he/she is done. The student is then asked to do the original problem again.

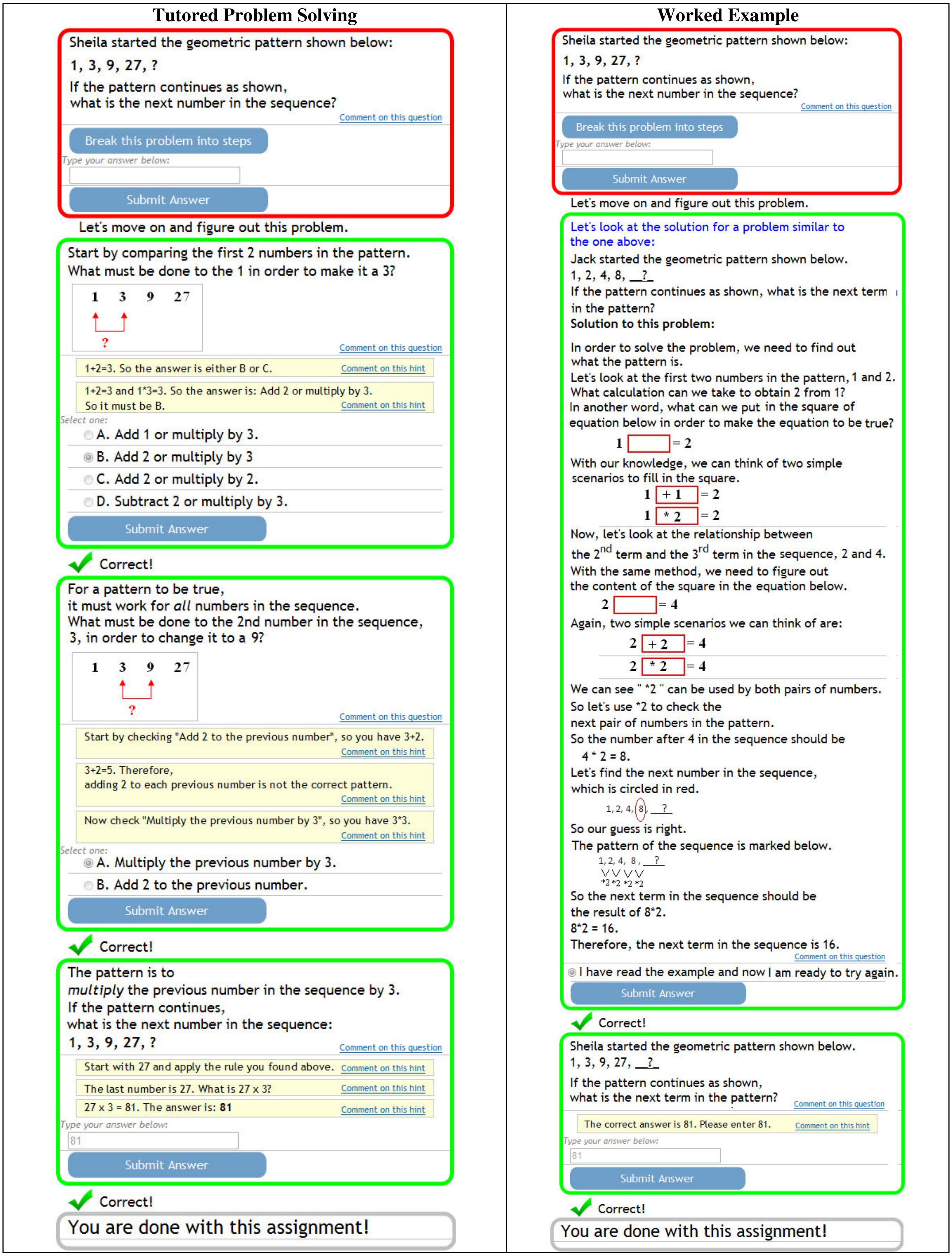

Figure 2: Screen shots of the tutored problem solving condition on the left and the worked example condition on the right.

During the experiment, teachers introduced the problem sets as a regular assignment. As such, students were not aware of the randomized controlled experiment. They were neither briefed about the problem set structure nor the number of ASSISTments in a problem set. Thus, students might not have been aware that they were taking a pre-test until they submitted an answer, as we tell them that the question they answered was a pre-test only after answer submission.

We do not distinguish the experiment section from the post-test with any specific instruction or notice like we do in the pre-test. The only way a student can know that they are in the post-test is if they realized that the pre-test question has been repeated. It should be noted that there is a possibility that some students were not exposed to either of the conditions since conditions are introduced only when a student makes a mistake in the first response. If students answered all of the ASSISTments in a problem set correctly in their first attempt then they would not have been exposed to any of the conditions and their performance on that problem set were not included in the study.

Results

Our experiment used a repeated measures design where students participated in a different number of experiments, and each time the student started an experiment, he/she was randomly assigned to one of the two conditions. Each problem set attempted by the students is recorded and defined as a data point. For the analysis, we only considered the students who had completed at least one problem set in both of the conditions and ignored all other students who were exposed to only one condition. All incomplete data points were also ignored. Moreover, we removed all of the data points where the student correctly answered both the pre-test and the post-test questions, as we assumed the student had mastered that material. Since repeated measure design suffers from ordering effects, we relied on the random assignment of conditions as a control for that effect.

Out of a total of 186 participants, 166 students completed at least one problem set and we had a total of 866 data points. We then ignored all of the data points where both pre-test and post-test answers were correct. Of the remaining 409 data points from 127 students, we again ignored all of the data points from students who completed only one of the two conditions. We then had a total of 318 data points from 68 students who participated in both tutoring conditions. So this means each of the 68 students participated in at least one problem set where they were given tutored problem solving and at least one problem set where they were given worked examples.

For each student, the average learning gain from tutored problem solving and the average learning gain from worked examples were calculated. Learning gain for a data point (problem set) was defined to be the post-test score minus the pre-test score. Average learning gain for the tutored problem solving condition was defined to be the average of the learning gains for all of the problem sets that the student did when they were assigned to the tutored problem solving condition. Similarly, the average learning gain for worked examples was the average of the gains for all of the problem sets that the students did when they were assigned to the worked examples condition. There was no need to check if both groups were balanced at pre-test since our experiment was a repeated design and each student participated in both conditions.

There was a significant effect for condition with tutored problem solving receiving higher gain scores than worked examples (34% average gain vs. 13% average gain), $t(67) = 2.38$, $p = 0.02$.

To determine whether there was an aptitude-treatment interaction we calculated a student math proficiency score using an Item Response Theory (IRT) model which takes into account difficulty of ASSISTments and how students performed on ASSISTments throughout the school year.

We did not have IRT scores for five students, so this analysis was done on data from 63 students. We did not find a significant difference based on math proficiency ($F(1, 61) = .158$, $p = 0.69$). While both high proficiency and low proficiency students learned more from the tutored problem solving condition than from the worked example condition, low proficiency students appeared to learn very little from the worked examples. High proficiency students had a mean gain of 47% with tutored problem solving and 22% with worked examples ($t(29) = 1.599$, $p = 0.12$). Low proficiency students had a mean gain of 20% with tutored problem solving and 3% with worked examples ($t(32) = 1.404$, $p = 0.17$). These results are shown in Figure 3.

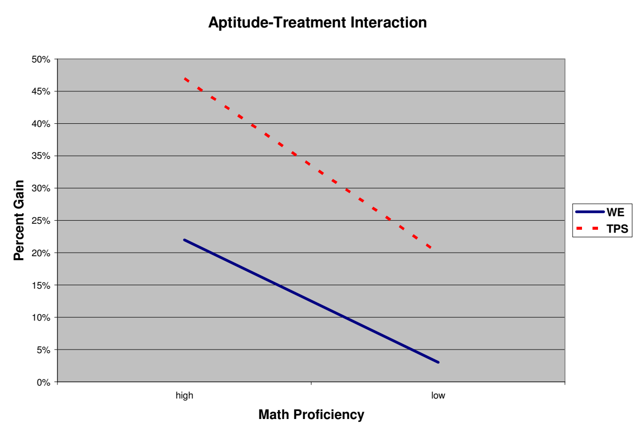

Figure 3: There was no significant aptitude-treatment interaction.

Because the tutored problem solving is more interactive, it does consume more time, so we have a time on task confound. Tutored problem solving (M = 244 seconds) took significantly more time on average than worked examples (M = 166 seconds), ($t(66) = 2.93$, $p = 0.002$).

Discussion and Contributions

Our study compared the effectiveness of tutored problem solving versus worked examples when used as feedback while paying attention to students' math proficiency. Students participated in the study in a classroom environment and the problems were presented as regular classroom assignments. Our results indicate that tutored problem solving is significantly better than worked examples in terms of the average gain of students in each condition. We did not find an aptitude-treatment interaction.

Our study differed from previous studies in that we compared worked examples to tutored problem solving rather than untutored problem solving. We also differ in that we presented worked examples as feedback after students unsuccessfully attempted to solve a problem rather than presenting them before they attempted problem solving. Our study is lacking in that we did not do a transfer or retention test. We also had a time-on-task confound since tutored problem solving takes significantly longer than worked examples.

Our results are different from what the existing literature would expect. We speculate that many studies that have found positive results for worked examples were done in research lab settings, where an adult lab attendant provided the extra focusing attention that a classroom environment does not provide. Perhaps in the classroom setting, the more interactive tutored problem solving condition was superior due to the fact that the higher interactivity level required from tutored problem solving better engages students' focus. This theory would suggest that students with greater focus might yield results that would be different and more in line with the current literature.

Salden et al. [15] thought of their results as an instance of the "Assistance Dilemma" coined by Koedinger and Aleven [9] which studies the dilemma of when to give assistance to students versus when to withhold information in an attempt to get students to generate information on their own. The Assistance Dilemma would consider worked examples to be "high assistance" while tutored problem-solving to be "low assistance". However, this does not seem to consider that these may be seen differently by students depending on how well-focused they are. For instance, Chi, Bassok, Lewis, Reimann, & Glaser [4] found a difference in the way that students used worked examples based on their proficiency in problem-solving: "… we find that the Good students use the examples in a very different way from the Poor students. In general, Good students, during problem solving, use the examples for a specific reference, whereas Poor students reread them as if to search for a solution." Recently we [12] found that students who received worked-out *solutions* to problems rather than tutored problem-solving learned more only if they were above average students. Below average students did better with tutored problem-solving. (We believe that our use of worked-out *solutions* is similar to worked examples in that they do not withhold information.)

We think our theory can explain the current results in this area. In particular, we speculate that the students in the recent Salden et al. study [15] might have been just the right type of well focused students that could benefit from reading worked examples. However, if you want to help the less focused student then tutored problem solving may be superior.

This conclusion is reasonable since these two conditions have different degrees of interactivity. In the worked examples condition, a student is shown a completely solved example problem which is similar to the main problem. The student is only one click away from answering the original problem again. In contrast, the tutored problem solving condition asks several subsequent questions pertaining to the main problem, all of which have to be completed before returning to the main problem again. For most students, it is reasonable to assume that answering questions frequently keeps them more focused than reading off of a screen.

Several studies in the literature have found evidence of the benefit of greater interaction with human tutors. Comparing Socratic and didactic tutoring strategies, Core, Moore and Zinn [5] found that the more interactive (based on words produced by students) Socratic tutorial dialogs correlated more with learning. Chi, Siler, Jeong, Yamauchi and Hausmann [3] found that students who engaged in a more interactive style of human tutoring were "able to transfer their knowledge better than the students in the didactic style of tutoring." Evens and Michaels [6] compared expert human tutoring to reading a text book with the same material and found that the tutored students got significantly higher scores on a post-test. Results that support greater interaction have also been found in studies of e-learning systems [7, 19].

It is possible that our results can be explained by cognitive load theory: perhaps the tutored problem solving reduces cognitive load even more than worked examples as students are walked through problems step by step and subgoals are set for them. There may be a tradeoff in that students may lose the big picture by working on pieces of a problem at a time and are not asked to induce principles, but subgoal learning has been found to help guide problem solving by helping learners focus on the steps [2]. On the other hand, McLaren et al. [11] found little difference in learning gains between tutored problem solving alone and tutored problem solving interleaved with worked examples and could not adequately explain their results with cognitive load theory.

We believe our results are important because it raises the question about whether worked examples are always a better thing to do before problem solving for all students as other studies found [15, 16]. Obviously, worked examples are useful and even necessary when a student is first introduced to a concept. But how long does this benefit last?

A logical follow up study would be if we controlled for the level of interactivity in the two conditions by asking students questions about the worked example they read.

Another logical study to conduct will be to see if when time on task is controlled for, if students with relatively low motivation can benefit from reading worked examples.

In conclusion, the results of our study show that worked examples alone are not more effective than tutored problem solving. The key may be in how interactive the tutoring strategy is.

Acknowledgments

We would like to thank all of the people associated with creating the ASSISTment system, listed at www.ASSISTment.org, and acknowledge funding from the US Department of Education, the National Science Foundation, the Office of Naval Research and the Spencer Foundation. The first author is funded by a GK12 NSF Fellowship. All of the opinions expressed in this paper are solely of the authors and not of our funding organizations.

References

1. Anderson, J. R., Corbett, A., Koedinger, K., & Pelletier, R. (1995) Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences* 4(2), 167-207.
2. Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127* (4), 355-376.
3. Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from tutoring. *Cognitive Science*, 25:471-533.
4. 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.
5. Core, M. G., Moore, J. D., and Zinn, C. The Role of Initiative in Tutorial Dialogue, the 10th Conference of the European Chapter of the Association for Computational Linguistics.
6. Evens, M.W., & Michael, J.A. (2006). *One-on-one tutoring by humans and machines*. Mahway, NJ: Lawrence Erlbaum Associates.
7. Graesser, A. C., Moreno, K., Marineau, J., Adcock, A., Olney, A., & Person, N. (2003). AutoTutor improves deep learning of computer literacy: Is it the dialog or the talking head? In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Proceedings of Artificial Intelligence in Education* (pp. 47–54). Amsterdam: IOS.
8. Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. Journal of Educational Psychology, 93(3), 579–588.
9. Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review, 19*, 239-264.
10. Larkin, J., McDermott, J., Simon, D. & Simon, H. (1980) Expert and novice performance in solving physics problems. *Science, 208,* 1335-1342.
11. McLaren, B.M., Lim, S., & Koedinger, K.R. (2008). When and How Often Should Worked Examples be Given to Students? New Results and a Summary of the Current State of Research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2176-2181). Austin, TX: Cognitive Science Society.
12. Razzaq, L., Heffernan, N. T., Lindeman, R. W. (2007) What level of tutor interaction is best? In Luckin & Koedinger (Eds) Proc. of the 13th Conf. on Artificial Intelligence in Education. IOS Press. pp. 222-229.
13. Razzaq, L., Heffernan, N. T., Koedinger, K. R., Feng, M., Nuzzo-Jones, G., Junker, B. Macasek, M. A., Rasmussen, K. P., Turner, T. E., & Walonoski, J. A. (2007). Blending Assessment and Instructional Assistance. In Nadia Nedjah, Luiza deMacedo Mourelle, Mario Neto Borges and Nival Nunesde Almeida (Eds). Intelligent Educational Machines within the Intelligent Systems Engineering Book Series. 23-49 Springer Berlin / Heidelberg.
14. Ringenberg, M. A., VanLehn, K. Scaffolding Problem Solving with Annotated, Worked-Out Examples to Promote Deep Learning. Intelligent Tutoring Systems 2006: 625-634.
15. Salden, R., Aleven, V., Renkl, A., & Schwonke, R. (2008). Worked examples and tutored problem solving: redundant or synergistic forms of support? In C. Schunn (Ed.) Proceedings of the 30th Annual Meeting of the Cognitive Science Society, CogSci 2008 (pp. 659-664). New York, NY: Lawrence Earlbaum.
16. Schwonke, R., Wittwer, J., Aleven, V., Salden, R., Krieg, C., & Renkl, A. (2007). Can tutored problem solving benefit from faded worked-out examples? In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of EuroCogSci 07. The European Cognitive Science Conference 2007* (pp. 59-64). New York, NY: Erlbaum.
17. Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. Cognition and Instruction, 2, 59–89.
18. VanLehn, K., Lynch, C., Schulze, K. Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D.,Weinstein, A., & Wintersgill, M. (2005) The Andes physics tutoring system: Lessons Learned. In *International Journal of Artificial Intelligence and Education,* 15 (3), 1- 47.
19. VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. When is reading just as effective as one-on-one interactive human tutoring? In Proceedings of the 27th Annual Meeting of the Cognitive Science Society (pp. 2259-2264). Mahwah, NJ: Erlbaum. (2005) 2259-2264.
20. Warnakulasooriya, R. & Pritchard, D. E. (2005) Learning and problem-solving transfer between physics problems using a web-based homework tutor. *EdMedia – Conference on Educational Multimedia, Hypermedia & Telecommunications.* Vol. 2005, 2976-2983.