Measuring Learning Progress via Self-Explanations versus Problem Solving - A Suggestion for Optimizing Adaptation in Intelligent Tutoring Systems

Christine Otieno (otieno@psychologie.uni-freiburg.de)
Department of Educational and Developmental Psychology, University of Freiburg
Engelbergerstr. 41, D-79085 Freiburg, Germany

Rolf Schwonke (schwonke@psychologie.uni-freiburg.de)
Department of Educational and Developmental Psychology, University of Freiburg
Engelbergerstr. 41, D-79085 Freiburg, Germany

Alexander Renkl (renkl@psychologie.uni-freiburg.de)
Department of Educational and Developmental Psychology, University of Freiburg
Engelbergerstr. 41, D-79085 Freiburg, Germany

Vincent Aleven (aleven@cs.cmu.edu)
Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213 USA

Ron Salden (rons@cs.cmu.edu)
Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213 USA

Abstract

Prior studies have shown that learning by problem solving in an intelligent tutoring system such as the Cognitive Tutor can be even more effective when worked examples are added in comparison to problem solving alone. Introducing self-explanation prompts additionally improves learning. Furthermore, recent findings indicate that fading out worked examples according to students' performance during learning (i.e., adaptive fading) is even more beneficial than fading worked examples in a predefined procedure (i.e., fixed fading). In this contribution we investigate the relationship between potential indicators for learning progress, which can be used for adapting fading and, thereby, fostering learning outcome. We found a stronger relationship of learning outcomes to self-explanation performance than to problem-solving performance during learning. Additionally, self-explanation performance is a stronger predictor for learning outcome than prior knowledge. Hence, adaptation, not only of the example fading procedure but potentially of other aspects of student learning (e.g., individualized problem selection) might better be based on self-explanation performance and not, or at least not only, on problem-solving performance, as it is typical of intelligent tutoring systems.

Keywords: Scaffolding, Worked Examples, Intelligent Tutoring Systems, Adaptive Fading

Introduction

Nowadays individualized instruction is a catchphrase that is becoming more and more important. Cognitive Tutors and other intelligent tutoring systems have proven to be very effective in supporting individual students' learning in a variety of domains such as mathematics or genetics (for an overview, see Koedinger & Corbett, 2006). Cognitive Tutors are used in more than 2600 schools across the United States as part of the regular curriculum. Based on an online assessment of students' learning, Cognitive Tutors provide individualized support for guided learning by problem solving (doing). Specifically, the Tutor selects appropriate problems, gives just-in-time feedback, and provides hints.

Introducing self-explanation prompts to the Cognitive Tutor made the Tutor even more effective (Aleven & Koedinger, 2002). However, from a cognitive load perspective the introduction of self-explanation activities in addition to problem solving places fairly high demands on students' limited cognitive capacity, particularly in the early stages of skill acquisition (e.g., Sweller, van Merriënboer, & Paas, 1998), notwithstanding the load reducing aspects of Cognitive Tutors, such as making subgoals and intermediate steps explicit. The additional load posed by self-explanations can be reduced by scaffolding the learning process with worked examples (e.g., Salden, Aleven, Renkl, & Schwonke, 2009). Meanwhile, there is ample empirical evidence showing that learning from worked examples leads to superior learning outcomes as compared to problem solving (for an overview, see Renkl, 2011).

Although studying worked examples has proven to be effective, this is true only during early stages of skill acquisition (e.g., Kalyuga, Chandler, Tuovinen, & Sweller, 2001). During that phase, scaffolding with worked examples reduces the cognitive demands of problem solving and allows the learner to focus on understanding domain principles. As expertise increases, worked examples not only lose their effectiveness but can even impede learning progress (*expertise reversal effect*; Kalyuga, Ayres, Chandler, & Sweller, 2003). The more students know about a subject matter the more emphasis should be put on problem-solving activities which lead to an increase of speed and higher accuracy of performance (Renkl & Atkinson, 2003). Therefore, Renkl and Atkinson (2003) proposed a fading procedure in which problem-solving elements are successively integrated into example study until the students are able to solve problems on their own, that is, scaffolding is reduced according to students´ expertise.

In a number of experiments, Renkl and colleagues provided empirical evidence for the effectiveness of a smooth transition from example study to problem solving (e.g., Atkinson, Renkl, & Merrill, 2003; Renkl, Atkinson, & Große, 2004). The specifics of the sequence in which worked examples are faded are crucial for the optimization of learning. Although these studies indicate that worked examples faded in a fixed procedure were superior to example-problem pairs, similar like in classical research about scaffolding (e.g. Wood, Bruner, & Ross, 1976), fading worked examples adaptively to the individual learner's progress should be even more successful. By assessing the learning progress one can avoid the negative effects of worked examples, also known as the reverse worked example effect (Kalyuga et al., 2001). The Cognitive Tutor with its online assessment provides an appropriate framework for fading worked examples adaptively.

Evidence for the effectiveness of adaptively fading worked examples was first found in one of our previous experiments (Salden et al., 2009). In this laboratory study, we compared three conditions: a problem-solving condition, a fixed-fading condition, and an adaptive-fading condition, in order to test effects of faded worked examples over problem-solving alone and adaptive fading over fixed fading of worked examples (see also Method section in this paper). As expected, Salden et al. (2009) found a monotonic trend of adaptive fading over fixed fading over problem solving. Effects were found in both posttest ($Z = 2.03$, $p < .05$) and delayed posttest ($Z = 1.83$, $p < .05$). Additionally, contrasts calculated to compare the adaptive-fading condition with the non-adaptive conditions (fixed fading and problem solving) revealed a significant superiority of the adaptive-fading condition in both immediate posttest ($t(54) = 1.74$, $p < .05$, $d = .49$) and delayed posttest ($t(49) = 2.04$, $p < .05$, $d = .59$). These findings could be largely replicated in a field experiment (Salden et al., 2009). Taken together, these results indicate that not only are faded examples superior to example-problem pairs, as already found in earlier studies (e.g., Schwonke, Renkl, Krieg, Wittwer, Aleven, & Salden, 2009), but also adapting the fading procedure according to students' performance is superior to a fixed sequence.

Typically, Cognitive Tutors adapt instruction based on the learner's problem-solving performance (Corbett, McLaughlin, & Scarpinatto, 2000). Unlike this widely used approach, the adaptation (here: of fading) in our experiment could not be based on problem-solving performance while working in the Cognitive Tutor, because problem-solving steps were worked-out in the beginning. Hence, we used self-explanation performance, that is, a type of meta-cognitive performance (Aleven & Koedinger, 2002), for adaptation. Against this background, the questions arise whether it is sensible at all to rely on self-explanation performance or whether this might be even the better indicator for learning progress. The finding of Salden et al. (2009) on the superiority of adaptive fading suggests that self-explanation performance is a sensible indicator for learning progress that can be used for adaptation, even if these self-explanations are prompted and supported by menus. However, in order to gain deeper insight in the potential of self-explanation performance as an indicator for adaptation and in the potential of different indicators, we performed a re-analysis of our laboratory study.

We assumed, that students who have difficulties in gaining deeper understanding make more mistakes while working with the Tutor (e.g., Aleven, McLaren, & Koedinger, 2006). Higher proportions of incorrect entries for both numerical entries (*answers)* and self-explanation activities (*reasons*) should therefore be associated with inferior learning outcomes (in terms of transfer performance). This (negative) relationship should be especially strong for self-explanation (i.e., reason) steps as we assume that they reflect a deeper understanding. Therefore, self-explanation performance in addition to the traditionally used problem-solving performance should be predictive of learning outcomes. More specifically, we addressed the following hypotheses:

(1) There is a negative relationship between performance (i.e., incorrect entries) on problem-solving (i.e., answer) and self-explanation (i.e., reason) steps while working with the Tutor and learning outcomes.
(2) The negative relationship is stronger for performance on self-explanations steps.
(3) Performance on problem-solving and self-explanation steps is both predictive of learning outcomes.
(4) Performance on self-explanations steps is a predictor of learning outcomes, beyond the predictive power of performance on problem-solving steps and prior knowledge.

Method

Sample and Design

We recruited 57 students (19 in 9th grade and 38 in 10th grade) from a German "Realschule", which is equivalent to an American high school. The participants (age: $M = 15.63$, $SD = .84$) were randomly assigned to one of the three conditions with 19 participants each. In two conditions students were given worked examples for problem-solving (i.e., answer) steps which were either faded out according to a fixed procedure (*fixed fading condition*) or according to the student's individual skill level and self-explanation performance (*adaptive fading condition*). The third condition did not receive any worked examples (*problem condition*) and served as a control condition. Students in all conditions had to provide prompted self-explanations (i.e., reasons) for all problem-solving steps and all students had to solve at least some problem steps (Aleven & Koedinger, 2002). As the aim of our reanalysis was to investigate relationships between performance on problem and reason steps while working with the Tutor and performance on posttest independent of condition, the following results refer to all 57 participants of the study.

Learning Environment – The Cognitive Tutor

In order to provide feedback and adapt to students' skill acquisition, Cognitive Tutors are based on so called production rule models. Different *production rules* for *knowledge components* can be learned independently. In the present case, a knowledge component represents specific ways of applying principles, for example, angle addition, that are to be learned by the student.

The assistance in the Geometry Cognitive Tutor is based on two algorithms: *model tracing* and *knowledge tracing* both of which are grounded in the idea of knowledge components in the production rule model. This model enables the tutor to simulate the problem solving process, to decide whether a student's action is right or wrong and to provide intelligent feedback (model tracing) as well as to estimate the student's learning progress on the level of knowledge components (knowledge tracing; Koedinger & Corbett, 2006). On this basis, the Cognitive Tutor can adapt the assistance given to students to their learning progress. Hence, we were also able to fade out worked examples adaptively in the adaptive fading condition. The type of problems that were presented in our study was held constant across conditions.

Learning Materials Students were asked to work on fifteen problems in a Cognitive Tutor lesson on geometry, together covering four geometry principles. The first eight problems required the application of only one geometry principle. The last seven problems combined different principles and were therefore more complex. In the *problem condition*, solving a problem required students (a) to enter numerical values such as the measure of an angle (i.e., the answer) and (b) to self-explain each given numerical answer (i.e., the reason). The self-explanation consisted of entering the name of the principle applied into a text entry field. The principle could be entered either by typing the name of the relevant principle or by selecting the principle from a glossary that contained a list of all principles used in the unit. For example, if angles AB and AC are complementary angles and the measure of angle AB is 60 degrees, then the measure of angle AC is 30 degrees, because the sum of the measures of complementary angles is equal to 90 degrees. The student would be required to either enter "90-60" or "30" on the answer step and "complementary angles" on the reason step.

In the two *example conditions*, students were asked to study a sequence of worked steps corresponding to the answer steps in the problem condition. Worked examples provided the numerical solutions of a problem step and necessary calculations. Students were then asked to provide a reason for the answers provided by the worked examples. The worked examples were gradually faded out according to either a fixed fading scheme or adaptively according to students' performance on self-explanation steps. Self-explanation activities were held constant across the three experimental conditions.

Instruments

Pretest The pretest was implemented in the Cognitive Tutor and consisted of four geometry problems related to the lessons taught later during the learning phase with the program. All Cognitive Tutor help facilities (e.g., hints) were disabled during pretest. All participants completed the same pretest.

Posttest A posttest that consisted of the same problems as the pretest was implemented in the Tutor. Additionally, all participants were asked to complete a paper-pencil test immediately after working with the Tutor and one week later (delayed posttest). Both posttests were identical. The items in the paper-pencil tests differentiated between near-transfer and far-transfer problems (four items for near transfer and four items for far transfer). Near-transfer problems were isomorphic to the problems in the Tutor; far-transfer problems were structurally different but based on the same concepts. As in the example shown in Figure 1, students were given a structurally similar figure like in the Tutor for near-transfer items. They were then asked (in this example) to calculate angle IGT and angle TGH. Figure 2 shows an example for a far-transfer item. In this item students were given a cover story of a sailor who is navigating by the stars, in this case, the Southern Cross. Then they were asked to calculate angle CXD given that angle AXD is 45 degrees to find out in which angle the destination island was to the sailor's ship.

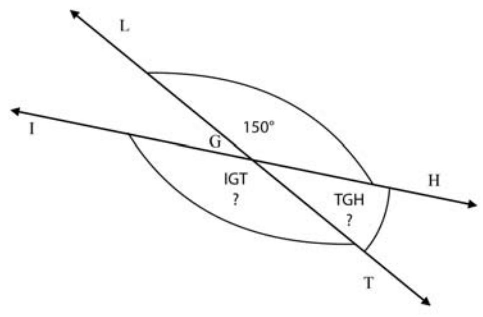

Figure 1: Example for a Near-Transfer Problem

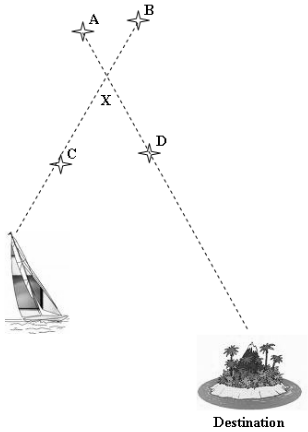

Figure 2: Example for a Far-Transfer Problem

Procedure

The experimental sessions lasted, on average, 90 minutes and were divided into three parts: pretest and introduction, tutoring, and posttest. First, students´ general prior knowledge was assessed by their mathematics grade together with some additional demographic data such as age and gender. Then they received a brief introduction on how to use the Cognitive Tutor followed by a short pretest implemented in the Tutor measuring their prior knowledge. After completing this pretest, students read an instructional text providing information about the rules and principles that were later addressed in the Cognitive Tutor. In the tutoring part, students worked with their respective version of the Cognitive Tutor. This learning phase was followed by the posttests.

Results

Table 1: H1 & 2: Relationship of Performance on Answer and Reason Steps and Learning Outcomes

		Answers	Reasons	
		r	*r*	N
Posttest	Near Transfer	-.34*	-.65***	57
	Far Transfer	-.18	-.48***	57
Delayed Posttest	Near Transfer	-.16	-.41**	52
	Far Transfer	-.12	-.49***	52

Note. * p < .05, ** p < .01, and *** p < .001 (two-tailed).

To test hypotheses 1 and 2 we determined Pearson´s correlations between the proportion of incorrect answers in relation to all answer steps executed (i.e., problem-solving) as well as the proportion of incorrect reasons (i.e., self-explanation) on the one hand and immediate as well as delayed posttest scores on the other hand (Table 1). Posttest scores reflect the percentage of points students received for their posttest of the total of points to be achieved. The performance on reason steps was significantly and substantially related to scores on near and far transfers items on both immediate and delayed posttest. In contrast, the performance on answer steps was only reliably related to scores on near transfer items in the immediate posttest. In fact, the relationships to learning outcomes were significantly stronger for reason steps than for answer steps as corresponding comparisons (Olkin) shows, z (near transfer) = 2.69, z (far transfer) = 2.29, z (delayed near transfer) = 1.87, z (delayed far transfer) = 2.81 (with z_{crit} = 1.65 for a one-tailed significance test with α= 5%). In summary, the pattern of results only partly confirmed hypotheses 1: Performance on reason steps (i.e., self-explanation), as indicator of deep understanding, was significantly related to posttest scores (i.e., learning outcomes), while performance on answer steps (i.e., problem solving) were not significantly related to posttest scores (except for near transfer on immediate posttest). Hypothesis 2 was confirmed: The negative relationship between performance on reason steps and learning outcome was significantly stronger than that for performance on answer steps and learning outcome.

Although, the performance on answer steps and on reason steps differed substantially in their predictive power with respect to the posttest measures, we found a medium correlation between them (r = .45, p < .001). This association can be expected because answer steps and reason steps are not independent but rather measure understanding on different levels. Moreover, performance on reason steps might be influenced by the Tutor's support received on the respective answer step.

Table 2: H3: Performance on Reason Steps as Predictor for Learning Outcomes (Final Regression Model)

			B	*SE B*	*β*
Posttest	Near Transfer	Reasons	-.24	.03	-.73***
	Far Transfer	Reasons	-.18	.04	-.50***
Delayed Posttest	Near Transfer	Reasons	-.12	.04	-.41**
	Far Transfer	Reasons	-.19	.05	-.49***

Note. Posttest, Near Transfer: R^2 = .54; Posttest, Far Transfer: R^2 = .25; Delayed Posttest, Near Transfer: R^2 = .17; Delayed Posttest, Far Transfer: R^2 = .24.
** p < .01 and *** p < .001.

To test hypotheses 3 and to decide if problem-solving or self-explanation activities or both in combination are presumably best for adapting support in intelligent tutoring systems, we calculated a stepwise linear regression with performance on reason and answer steps as predictors. As the correlations from Table 1 suggest, the predictive power of error rate on answer steps (i.e., problem-solving) was very low. Accordingly, regression models including only significant predictors omitted error rate on answer steps (Table 2). Hence, Hypothesis 3 was not confirmed, that is, only performance on reason steps but not on answer steps had significant predictive power for learning outcome.

Table 3: H4: Prior Knowledge and Performance on Answer and Reason Steps as Predictors for Learning Outcomes

			B	*SE B*	*β*
Posttest	Near Transfer	Step 1			
		Reasons	-.20	.03	-.65***
		Step 2			
		Reasons	-.18	.03	-.59***
		Math. Grade	-.04	.02	-.26*
	Far Transfer	Step 1			
		Reasons	-.17	.05	-.44**
		Step 2			
		Reasons	-.14	.05	-.37**
		Math. Grade	-.05	.02	-.27*
Delayed Posttest	Near Transfer	Step 1			
		Math. Grade	-.08	.02	-.53***
	Far Transfer	Step 1			
		Reasons	-.24	.05	-.55***
		Step 2			
		Reasons	-.21	.05	-.49***
		Math. Grade	-.05	.02	-.27*

Note. Posttest, Near Transfer: $R^2 = .42$ for Step 1, $\Delta R^2 = .06$ for Step 2 ($p < .05$); Posttest, Far Transfer: $R^2 = .19$ for Step 1, $\Delta R^2 = .07$ for Step 2 ($p < .05$); Delayed Posttest, Near Transfer: $R^2 = .28$; Delayed Posttest, Far Transfer: $R^2 = .30$ for Step 1, $\Delta R^2 = .07$ for Step 2 ($p < .05$).
* p < .05, ** p < .01, and *** p < .001.

To test Hypothesis 4 we calculated stepwise linear regressions with general prior knowledge measured by mathematics grade, specific prior knowledge measured by the pretest, performance on answer steps (i.e., problem-solving), and performance on reason steps (i.e., self-explanation) as potential predictors for learning outcome. Distributional assumptions were met by all dependent variables, that is, residuals could be assumed to be independent and distributed normally. Furthermore, heteroscedasticity could be assumed. (Multi-)Collinearity among predictors was not an issue, given tolerance values well above .2 and VIF values well below 10 (VIF values close to 1 for all predictors). Additionally, collinearity diagnostics showed that all predictors included in the models loaded highly on different dimensions.

Results indicate that specific prior knowledge as measured with the pretest did not yield additional explanatory power. However, general prior knowledge as measured with mathematics grade added predictive value to self-explanation activities in all models and even served as best sole predictor for near transfer in the delayed posttest (Table 3). These findings are in accordance with findings of strong influences of (general) prior knowledge on further learning (for an overview, see Dochy, Segers, & Buehl, 1999; Shapiro, 2004). On the whole, Hypothesis 4 is confirmed in that self-explanation performance has predictive power for learning outcomes beyond prior knowledge and problem-solving performance. Only in the case of the delayed near transfer, the hypothesis did not hold.

Discussion

Contrary to the widely used approach to base adaptation of instruction in intelligent tutoring systems on problem solving performance (i.e., answer steps), in the study by Salden et al. (2009) adaptation was based on self-explanation performance (i.e., reason steps). The superior learning outcomes of the adaptive fading condition shows that adapting on the basis of self-explanation is a feasible alternative. Our present findings indicate that it may even be the better alternative: Learning outcomes were better predicted by performance on reason steps (i.e., self-explanation) than by performance on answer steps (i.e., problem solving). In addition, regression models´ predictive power for learning outcome was not increased by including performance on answer steps. Again, given that traditionally adaptation is based on problem-solving activities this is a very "provocative" finding: Did we use only a sub-optimal indicator for students´ learning progress up until now?

Some students were able to write down mathematical values but failed to provide correct self-explanations. A similar discrepancy was observed by Siegler and Stern (1998) in strategy discovery and by Aleven, Koedinger, Sinclair, and Snyder (1998) for problem solving in the Geometry Tutor. It indicated that (in spite of the correct problem-solving performance) a full understanding of the problem-solving step is still lacking and still needs to be developed. Against this background, self-explanation performance might actually be a particularly sensitive indicator as to whether a student has actually understood a problem-solving step and should therefore be confronted with a higher problem-solving demand. In addition, the present findings suggest that a step should not be faded out before a "complete" understanding is achieved, that is, a student can also provide a correct self-explanation (i.e., reason) for a problem-solving step.

Our findings have also shown that general domain knowledge could be worth considering as a basis for initial adaptation. With respect to the finding that specific prior knowledge was less important, one should consider that the pretest used in this study was rather short. Using a more elaborate pretest might lead to different results. Additionally, one can assume that prior knowledge and self-explanation performance are not independent; prior knowledge can influence the quality of self-explanations. It might be argued that self-explanation performance is a more "proximal" indicator of specific knowledge than a pretest. Further studies have to clarify this issue.

The main message of this paper is that traditional adaptation procedures that are based solely on problem solving performance are presumably sub-optimal and that including self-explanation performance is likely to improve adaptation. However, the present findings need, without doubt, corroboration by further research that tests more directly the effects of different adaptation procedures. We suggest a comparison of at least three conditions in future studies: one in which online assessment and adaptation are based on self-explanation performance and one in which online assessment and adaption are based on problem-solving performance. A third group combining the two should be added to test if self-explanation as a single indicator is as good or even better compared to a combination of self-explanation and problem solving. Considering our results as well as those of many other studies, prior knowledge, especially general domain knowledge, should be taken into account as well.

Our results might also have important implications for classroom settings. Contrary to widely used methods of measuring students´ understanding by examining if they are able to solve problems correctly, it might be reasonable to test for students´ ability to explain their solutions rather than focusing on correct solution steps only.

Acknowledgments

This work was supported by the Pittsburgh Science of Learning Center which is funded by the National Science Foundation; award number SBE-0354420.

References

Aleven, V., McLaren, B. M., & Koedinger, K. R. (2006). Towards computer-based tutoring of help-seeking skills. In S. Karabenick & R. Newman (Eds.), *Help Seeking in Academic Settings: Goals, Groups, and Contexts* (pp. 259-296). Mahwah, NJ: Erlbaum.

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

Aleven, V., Koedinger, K. R., Sinclair, H. C., & Snyder, J. (1998). Combatting shallow learning in a tutor for geometry problem solving. In B. P. Goettl, H. M. Halff, C. L. Redfield, & V. J. Shute (Eds.), *Intelligent Tutoring Systems, Fourth International Conference, ITS '98* (pp. 364-373). Berlin: Springer Verlag.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology, 95*, 774-783.

Corbett, A., McLaughlin, M., & Scarpinatto, C. K. (2000). Modeling student knowledge: Cognitive Tutors in high school and college. *User Modeling and User-Adapted Interaction, 10*(2), 81-108.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69,* 145–186.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.

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.

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*. New York, NY: Cambridge University Press.

Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15-22.

Renkl, A. (2011). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 272-295). New York, NY: Routledge.

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How fading worked solution steps works – A cognitive load perspective. *Instructional Science, 32*, 59-82.

Salden, R., Aleven, V., Renkl, A., & Schwonke, R. (2009). Worked examples and tutored problem solving: Redundant or synergistic forms of support? *Topics of Cognitive Science, 1*, 203-213.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. J. C. M. (2009). The Worked-example Effect: Not an Artefact of Lousy Control Conditions. *Computers in Human Behavior, 25*, 258-266.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41*, 159-189.

Siegler, R. S. & Stern, E. (1998). Conscious and unconscious strategy discoveries: A micro-genetic analysis. *Journal of Experimental Psychology: General, 127,* 377-397.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring and problem solving. *Journal of Child Psychology and Psychiatry, 17,* 89-100.