Instr Sci
DOI 10.1007/s11251-009-9107-8

The expertise reversal effect and worked examples in tutored problem solving

Ron J. C. M. Salden · Vincent Aleven · Rolf Schwonke · Alexander Renkl

Received: 5 December 2008 / Accepted: 3 August 2009

Abstract Prior research has shown that tutored problem solving with intelligent software tutors is an effective instructional method, and that worked examples are an effective complement to this kind of tutored problem solving. The work on the expertise reversal effect suggests that it is desirable to tailor the fading of worked examples to individual students' growing expertise levels. One lab and one classroom experiment were conducted to investigate whether adaptively fading worked examples in a tutored problem-solving environment can lead to higher learning gains. Both studies compared a standard Cognitive Tutor with two example-enhanced versions, in which the fading of worked examples occurred either in a fixed manner or in a manner adaptive to individual students' understanding of the examples. Both experiments provide evidence of improved learning results from adaptive fading over fixed fading over problem solving. We discuss how to further optimize the fading procedure matching each individual student's changing knowledge level.

Keywords Cognitive tutor · Worked examples · Adaptive fading · Expertise reversal effect

Introduction

Research on worked examples in the context of Cognitive Load Theory (e.g., Sweller and Cooper 1985; Ward and Sweller 1990) has demonstrated that when students are presented

R. J. C. M. Salden (✉)
Human–Computer Interaction Institute, Carnegie Mellon University, 300 South Craig Street, Pittsburgh, PA 15213, USA
e-mail: rons@cs.cmu.edu

V. Aleven
Human–Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

R. Schwonke · A. Renkl
Psychological Institute Educational and Developmental Psychology, University of Freiburg, Engelbergerstr. 41, 79085 Freiburg, Germany

with example-problem pairs rather than problems only, they attain higher learning outcomes because their working memory capacity is not overloaded. Following this line of research, Kalyuga and colleagues (see Kalyuga 2007 for an overview) have established the expertise reversal effect which indicates that worked examples are more favorable in earlier stages of learning, while problem solving could be more effective in later stages. Worked examples reduce problem solving demands by providing worked-out solutions. Therefore, more of the learners' limited processing capacity (i.e., working memory capacity) can be devoted to understanding the domain principles and their application to the problem at hand (Renkl and Atkinson 2007).

A key implication of expertise reversal for instructional design is that worked-out steps should gradually be 'faded' from worked examples (complete solution is presented to the learner) to problems (learner must find the solution) (Atkinson et al. 2003; Renkl et al. 2002, 2004; Renkl and Atkinson 2007). Ideally, this transition should happen when the learner demonstrates understanding (e.g., by adequately linking worked-out steps to the underlying domain principles). Essentially, the instruction needs to be adapted to the learner's level of expertise by initially providing worked examples when the learner's working memory capacity is limited. Only when the learners gain understanding of the domain principles and can apply these in problem solving are the learners presented with actual problem-solving demands.

Determining the transition point is an interesting instructional design challenge and an instance of a broader "assistance dilemma," a fundamental choice that comes up in many instructional contexts (Koedinger and Aleven 2007): when, in an instructional sequence, is it more effective to provide assistance (e.g., example solutions), and when is it more effective to let the learner try to generate or construct solutions for themselves, without assistance from the system (or with lower levels of assistance)?

In the current research, we address this question by investigating the effectiveness of embedding example fading within an instructional approach that has been shown to be very successful in optimizing problem-solving performance, namely, tutored problem solving with Cognitive Tutor software (Anderson et al. 1995; Koedinger et al. 1997). These computer-based tutors provide individualized support for learning complex cognitive skills through problem-solving practice. They select appropriate problems to-be-solved, provide feedback and problem-solving hints, and assess each student's learning progress. Cognitive Tutors individualize the instruction by selecting problems based on a model of the students' present knowledge state that is constantly updated through a Bayesian process called "knowledge tracing" (Corbett and Anderson 1995). Adding self-explanation prompts to a Cognitive Tutor has been shown to increase the learners' understanding of how domain principles apply in problem solving (Aleven and Koedinger 2002; Roy and Chi 2005; cf. Van Lehn et al. 2005).

While there is an extensive body of research on worked examples, their use within a tutored problem-solving environment (e.g., a Cognitive Tutor) has not been tested until very recently (McLaren et al. 2007; Schwonke et al. 2007). Such an environment offers a significant amount of guidance to learners through step-by-step feedback and hints. As pointed out by Koedinger and Aleven (2007) "tutored problem solving" may be a substantially different control condition against which to measure possible beneficial effects of worked examples, compared to the untutored problem-solving conditions that were used as control conditions in the majority of previous research studies on worked examples and expertise reversal. On the other hand, Cognitive Tutors offer a number of advantages for research on worked examples and expertise reversal. Firstly, it is possible to capitalize on the capability of Cognitive Tutors to track each individual student's knowledge growth over time (Corbett and Anderson 1995). This knowledge assessment can be used as the basis for embedding an adaptive individualized example fading mechanism within a Cognitive Tutor. Secondly, Cognitive Tutors are "convenient" vehicles for classroom research: they are being used in over 2,600 schools across the United States. By contrast, many studies on worked examples are conducted in laboratory settings. Given the complementary advantages of lab and classroom studies, we conducted both of them in the research reported in this paper.

This research builds on earlier lab studies in which we investigated whether adding examples to tutored problem solving is more effective than tutored problem solving alone (Schwonke et al. 2007). In this approach, examples are added to tutored problem solving and are faded gradually according to a "fixed" fading scheme that is the same for all learners. Students self-explain the example steps, as well as their problem-solving steps using simple self-explanation menus to identify the geometry theorem that justifies each step (see Fig. 1). They also receive feedback on these self-explanations which has been shown to lead to better learning with a standard Cognitive Tutor (Aleven and Koedinger 2002). The results of these lab studies indicated that tutored problem solving combined with example fading lead to better transfer and reduced learning times compared to tutored problem solving alone.

As suggested by Kalyuga and Sweller (2004, 2005) and our own prior work (Schwonke et al. 2007), the fading of examples could be even more beneficial for learning if the rate at which the worked-out steps are faded would be adapted to the students' individual learning progress. Studying and self-explaining worked-out solution steps prepares learners to deal with subsequent problem-solving demands in a principle-based way. A learner who has not yet gained a basic understanding of a principle and of the way in which it is applied to

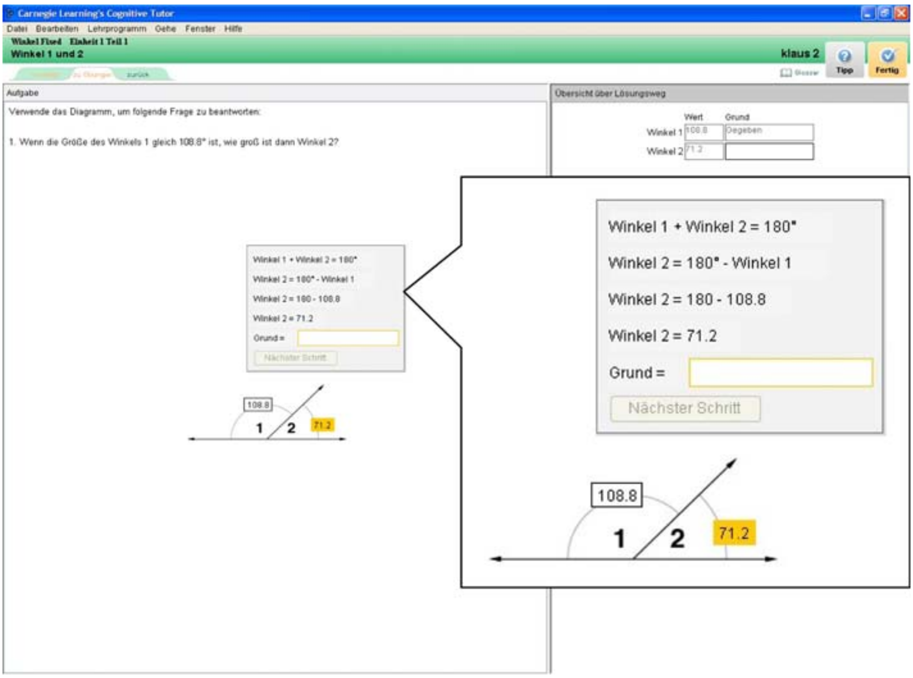

Fig. 1 Worked example in the German version of the Geometry Cognitive Tutor

solving problems should not be exposed to the corresponding problem-solving demands. Once the student shows a basic understanding of the principle and its application, s/he could go one step further and apply this knowledge to corresponding problem-solving steps. An adaptive fading procedure will make it more likely that students will get a fully worked step or a faded step dependent on their actual knowledge state.

Kalyuga and Sweller (2004, 2005) developed a Rapid Dynamic Assessment method which allowed them to update a model of each individual student's knowledge both prior and during the learning phase. Based on the student's prior knowledge level s/he was assigned to an appropriate level of faded worked examples. By periodically assessing the student's changing level of expertise, further adjustments were made to determine the optimal example fading point for each individual student. The Rapid Dynamic Assessment method was designed and validated using algebra and geometry materials (Kalyuga and Sweller 2004) and was shown to lead to higher knowledge and cognitive efficiency gains than a non-adaptive yoked control group (Kalyuga and Sweller 2005). In the yoked control group, each student was linked to the problems-examples trajectory of a corresponding student in the adaptive group. However, the students in the yoked control group were not assessed dynamically, and the level of faded worked examples was not based on their own learning progress. A restriction of this experiment was that the students in the yoked control group did not receive rapid dynamic assessments during the learning phase so that there could be different practice opportunities in both groups. It cannot be ruled out that the adaptivity effect found in this study is (in part) due to more practice.

We share the goal of adapting the transition from examples to problems to individual students' needs, but we address this question in the context of tutored problem solving with intelligent software tutors. Further, we base the transition not on rapid assessment of students' problem-solving ability, but on their ability to self-explain example steps. Specifically, as learners explain example steps in the manner described above, the tutor's knowledge-tracing mechanism keeps track of how well they appear to understand and interpret example steps in terms of problem-solving principles, separately for each principle targeted in the instruction. Once that understanding reaches a certain threshold, problem steps that involve the particular principle are "faded," meaning that they are presented as open problem steps for the student to solve.

The advantage of fading examples adaptively based on learner self-explanations rather than on Rapid Dynamic Assessment is that no separate assessment procedure is needed. Students' self-explanations that they are asked to generate for all steps in all examples and problems, can serve as a dynamic and rapid measure of student understanding. This assessment method is less intrusive, an advantage when operating in actual classrooms where students' attitudes towards any form of assessment might affect their motivation and performance. In addition, our adaptation procedure does not increase the number of practice opportunities (an advantage for setting up proper experimental control conditions). However, it is an open question whether a fading procedure based on the quality of students' self-explanations can be as effective as one based on the assessment of problem-solving skills, especially for the somewhat limited self-explanations expressed through a computer interface. It is not clear a priori whether these self-explanations are a rich enough source of data about student understanding or ability.

To investigate the instructional effectiveness of adapting worked-out examples in tutored problem solving based on the quality of individual students' self-explanations of worked-out steps, three experimental conditions were compared: (1) a problem solving condition that uses the standard Cognitive Tutor; (2) an example-enhanced Cognitive Tutor that fades worked-out steps in a fixed manner; and (3) an example-enhanced Cognitive

Tutor that fades worked-out steps adaptively for each individual learner. It was hypothesized that the adaptive fading procedure, combined with tutored problem solving will lead to better learning than a pure tutored problem-solving procedure and a fixed non-adaptive procedure for fading examples.

We conducted two experiments, both comparing these three experimental conditions, a lab experiment and an in vivo (i.e., classroom) experiment. Implementing and evaluating the same manipulations in both settings enables us to assess whether and how effects found in the lab setting transfer to the real-life environment. Such transfer cannot be taken for granted, given the many sources of variability in the classroom that are typically absent in the lab (e.g., distractions such as announcements over the intercom, students arriving late, off-task behavior, absenteeism, informal peer helping in ways that may or may not be effective etc.), and the fact that classroom studies often take place over longer periods of time. Thus, an ecologically valid investigation of the experimental manipulations and a "clean" lab investigation complement each other, and possible effects will have stronger implications.

Experiment 1: Lab study

Based on the results of Schwonke et al. (2007; fixed fading > problem solving) and of Kalyuga and Sweller (2004; adaptive fading > non-adaptive fading), we hypothesize a monotonic trend with respect to learning outcomes: adaptive fading > fixed fading > problem solving.

Method

Participants

For this study, 57 students (19 in 9th grade; 38 in 10th grade) were recruited from a German "Realschule," which is equivalent to an American high school. The participants (age $M = 15.63$, SD $= .84$) were randomly assigned to the three experimental conditions.

Materials

Cognitive Tutor. The experiment focused on a unit in the Geometry Cognitive Tutor that deals with the geometric properties of angles covering four theorems: angle addition, separate complementary angles, vertical angles, and linear pair. Every aspect (interface, problem statements, hints, and glossary) of the Cognitive Tutor was translated into German. We added new geometry problems covering the theorems of the selected unit because the fading procedure required problems that involve particular skill combinations. The problems were sequenced from simpler to more complex, with one-step problems presented first, followed by two-step problems, and eventually by three-step problems. More specifically, by "steps" we mean the application of a geometry theorem to find an unknown quantity. For instance, a one-step problem may involve application of the vertical angles theorem to find an unknown quantity (an angle measure); a two-step problem may involve the vertical angles theorem and the angle addition postulate; and a three-step problem may involve the rules for vertical angles, angle addition and linear pair of angles.

During the Cognitive Tutor training, the tutor provided its usual step-by-step guidance. That is, students received correctness feedback from the tutoring software after each step they performed. Furthermore, they could request hints at any point in time. For each step, several hint levels were available explaining which problem-solving principle applied, and how. The final hint level essentially stated the answer.

An example of a worked-out step for the linear pair ("Lineares Paar") theorem is shown in Fig. 1. The enlarged work area shows the worked-out steps and the self-explanation to be provided by the student. The value for the quantity sought in this step ("Winkel 2") is provided by the tutor. The student has to explain the step by indicating which theorem is used. To fill in this explanation (called "Grund") the student can either type the name of the theorem, or select the theorem from the tutor's online glossary of geometry knowledge. Figure 1 shows the "Glossar" hyperlink in the upper right corner that will open a glossary window in which students can browse relevant theorems and definitions (described and illustrated with simple examples).

In the Problem Solving condition, all steps of all problems were "pure problem solving," meaning that the students had to solve them. By contrast, in the Fixed Fading condition, as detailed in Table 1, students started out with fully worked-out examples, with example steps gradually being faded in subsequent problems until, in the last two problems, all steps were pure problem solving. More specifically, in problems P1 to P11, steps involving theorems T1 to T4 were worked-out (W) initially, and were systematically faded later (S for solving).

In the Adaptive Fading condition, the presentation of worked-out steps was the same as in the fixed fading condition up until the three-step problems (problems 7–11). Once students reached those problems, any step could be presented as either pure problem solving or worked-out step, depending on the student's performance in explaining worked-out steps in earlier problems that involved the same geometry theorem.

Fading decisions. The fading decisions were based on the tutor's estimates of each individual student's ability to produce valid explanations of worked-out steps. The tutor maintains estimates of the probability that the student is able to adequately explain a step (separately for each of the four theorems) using a Bayesian knowledge-tracing algorithm (Corbett and Anderson 1995). The estimates are updated each time the student explains a step involving the given geometry theorem; how much the estimate is increased or

Table 1 Instructional sequences used in the problem solving and fixed fading conditions

	Problem solving				Fixed fading			
	T1	T2	T3	T4	T1	T2	T 3	T 4
P1	S				W			
P2		S				W		
P3			S				W	
P4				S				W
P5	S		S		W		W	
P6		S		S		W		W
P7	S	S	S		W	W	S	
P8		S	S	S		S	S	W
P9	S	S		S	W	S		S
P10	S		S	S	S		S	S
P11	S	S	S		S	S	S	

T theorem, *P* problem, *S* solving, *W* worked-out decreased depends on whether the explanation was correct or not. The knowledge-tracing algorithm is a well-established method for student modeling in intelligent tutoring systems. It depends on four parameters for each skill (each theorem is considered to correspond to a separate skill): the probability that the student knows the skill coming in, the probability that the student will learn the skill as a result of applying it in a problem-solving step, and the probabilities that the student either guesses right or slips on a step involving the skill. In prior research, Cognitive Mastery Learning built on top of Bayesian Knowledge Tracing has been shown to significantly improve student learning (Corbett and Anderson 1995). Further, the estimates of skill mastery based on the Bayesian knowledge-tracing algorithm have been shown to accurately predict students' posttest scores (Corbett and Anderson 1995).

In the current project, in order to achieve effective fading of worked-out steps, the estimates of an individual student's mastery of each of the geometry theorems (i.e., the probability estimates coming out of the knowledge-tracing algorithm) were compared against two thresholds set at .7 and .5 respectively. Initially, when the estimated mastery level for the given skill/theorem is below the two thresholds, the tutor presents steps that involve the given skill (or geometry theorem) as worked-out steps. Each time the student explains one of these worked-out steps correctly, the tutor's mastery estimate will go up. Once the tutor's estimate of the student's level of understanding reaches the higher of the two thresholds (the .7 level), worked-out steps involving that theorem are faded. Accordingly, the steps in tutor problems that involve the skill/theorem are presented as open steps for the student to solve.

However, even after a student attains this level of understanding, s/he may make errors on subsequent steps that involve the same theorem, which will cause the tutor's estimate of mastery of that theorem to drop. If and when the estimate of skill mastery falls below the lower threshold (the .5 level), the tutor will again present the student with a worked-out step for the given theorem, until s/he reaches the higher threshold again. In this manner, the Adaptive Fading method is tailored to each individual student's evolving level of understanding. Typically, for any given skill/theorem, a learner will transition from examples to problems only once, but as described, there is the possibility of "regressing" back to examples. The fading points that result from this procedure are diagrammed in Fig. 2. The procedure worked in the exact same manner for each skill. That is, the fading points depended in the same way on the students' self-explanation performance for each skill, a result of setting the four knowledge-tracing parameters mentioned above to the same values for each skill. However, the adaptive fading procedure can be applied just as easily with skills for which the knowledge-tracing parameters have different values for each skill. Indeed, as discussed below, it is one of our recommendations for future studies.

Furthermore, it should be noted that the Cognitive Tutor's mastery learning mechanism was turned off during the lab study (Experiment 1) to obtain a clear-cut comparison between the three experimental conditions. As mentioned above, when this mechanism is on (as it is in the tutor's normal mode of operation), the tutor will let students finish a curriculum section only when they have mastered all skills targeted in a given curriculum section. It will select problems with unmastered skills until that criterion is met. By turning the mechanism off all students in the lab study completed the exact same problem set.

Posttest. The paper posttest consisted of 7 problems in total with 15 sub-steps each of which addressed a certain skill. Scoring was based on correctly answering the sub-steps, and the maximum score for the posttest was 85. The posttest problems consisted of three different types of tasks. The first task type constituted word problems from different real-world contexts with a different structure than that in the Tutor problems. The word

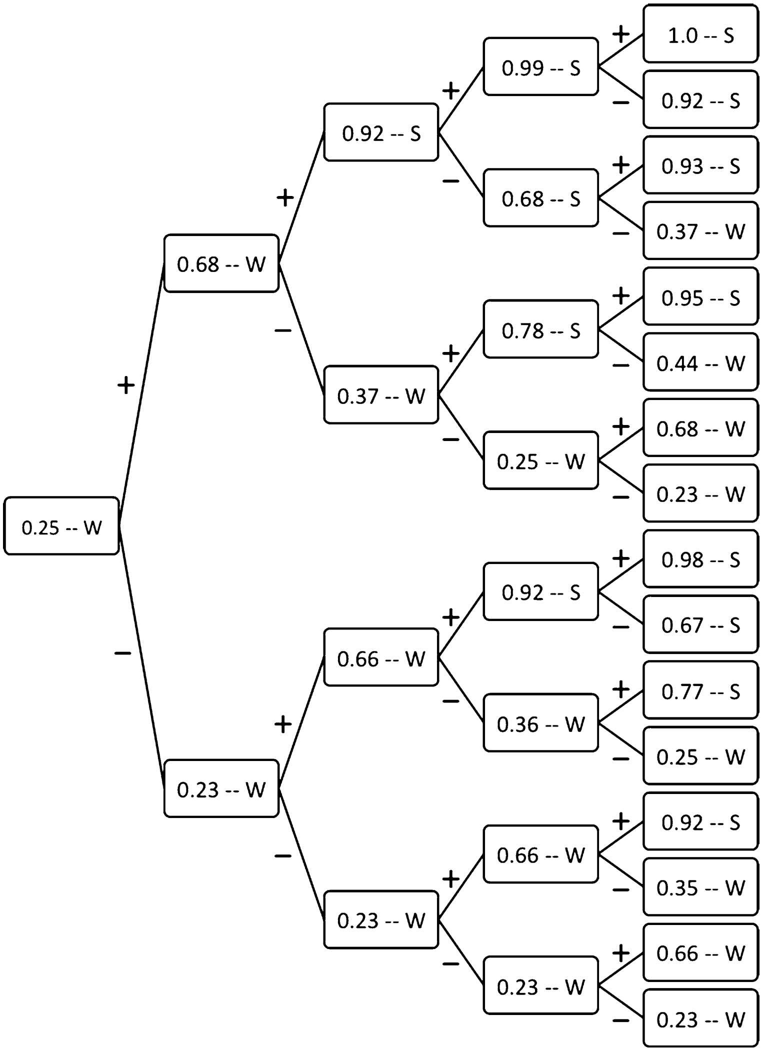

Fig. 2 Overview of the adaptive fading procedure based on the tutor's knowledge-tracing estimates of student knowledge. *Note*. The "+" and "−" on the links indicate whether the step was solved and/or explained correctly. The *number* in each *box* represents the tutor's estimate of the probability that the student understands the skill as a result of the sequence of correct and incorrect answers leading up to the *box*. The "W" or "S" in each *box* indicates whether (as a result of this estimate), the step would be worked out (W) or presented as an open step for the student to solve (S)

problems contained only a problem statement and a question asking students to find certain angles. Thus, the word problems (2 problems; 4 sub-steps) were open-ended whereas the Tutor problems had a clear step-by-step structure. In another task type, participants had to decide whether a given problem was solvable and, if it was, to provide the principles (2 problems; 5 sub-steps). In the third task type they had to generate real world examples for the given principles and to illustrate each example by a drawing (3 problems; 6 sub-steps). In other words, the posttest contained both procedural (first task type) and conceptual (second and third task type) knowledge items.

Procedure

The experiment consisted of two lab sessions. Since the students were unfamiliar with the Cognitive Tutor, they received paper-based instructions before using the tutor during the first lab session. Next, the students completed the actual Cognitive Tutor training (11 problems plus 1 warm-up problem) and the paper-based posttest. During the second session, which occurred 1 week later, a delayed paper-based posttest was administered which contained the same procedural and conceptual knowledge tasks as the immediate posttest. The students received 20 euro for their participation in the study.

Results

An alpha-level of .05 was used for all statistical analyses. Cohen's d was chosen as effect size measure—qualifying values <.3 as weak effect, values in the range between .3 and .5 as medium effect, and values >.5 as strong effect (see Cohen 1988).

Posttest performance

Using Jonckheere–Terpstra tests (one-tailed) to test the hypothesized monotonic trend (Adaptive Fading > Fixed Fading > Problem Solving), overall effects are found on the posttest ($Z = 2.03$, $P < .05$) and the delayed posttest ($Z = 1.83$, $P < .05$). The patterns of the means (immediate posttest: $M = 5.79$, SD = 3.44, for the problem solving condition; $M = 5.21$, SD = 3.20 for fixed fading condition; $M = 7.13$, SD = 3.37 for adaptive fading condition; delayed posttest: $M = 4.66$, SD = 3.70 for problem solving condition; $M = 4.50$, SD = 3.34 for fixed fading condition; $M = 6.67$, SD = 3.52 for adaptive fading condition) seem to indicate that the effect is mainly due to the superior performance of the adaptive condition. When contrasting the fixed fading condition with the problem solving condition no significant differences in posttest and delayed posttest performance were found (Fs < 1). However, when contrasting the adaptive fading condition with the other two conditions, the adaptive fading condition attained a higher performance on the immediate posttest ($t(54) = 1.74$, $P < .05$, $d = .49$) as well as on the delayed posttest ($t(49) = 2.04$, $P < .05$, $d = .59$). There were no differences in time spent on either of the posttests (Fs < 1).

Number of examples

In order to test whether the adaptive fading condition yielded a different total number of worked-out steps than the fixed fading condition, a t test was conducted. This test showed that the adaptive fading condition ($M = 8.42$, SD = 4.10) received significantly fewer worked-out steps than the fixed fading condition ($t(18.11) = 3.86$, $P < .01$, $d = 1.25$; $M = 12.05$, SD = .23). More specifically, when looking separately at the four theorems targeted in the Cognitive Tutor, the adaptive fading condition received significantly fewer worked steps than the fixed fading condition on *vertical angles* ($t(13) = 5.14$, $P < .0001$, $d = 1.94$) and *angle addition* ($t(18) = 4.23$, $P < .0001$, $d = 1.37$). The number of worked-out steps for each skill is shown in Table 2. There were no differences in time spent working in the Cognitive Tutor ($F < 1$).

Table 2 Means and standard deviations for the number of examples per theorem in faded conditions

	Fixed fading		Adaptive fading	
	M	SD	*M*	SD
Lab study				
Section 1				
Vertical angles***	3.00	.00	1.50	1.09
Linear pair	2.95	.23	3.58	2.27
Separate complementary angles	2.05	.23	1.80	1.52
Angle addition***	4.00	.00	2.32	1.73
In vivo study				
Section 1				
Vertical angles	3.00	.00	4.29	3.02
Linear pair*	2.88	.33	3.82	2.19
Separate complementary angles*	2.00	.00	3.00	1.37
Angle addition	3.76	.56	4.18	2.56
Section 2				
Angle subtraction	2.00	.00	2.42	1.73
Angle addition three angles	2.00	.00	2.25	.62
Separate supplementary angles	4.00	.00	3.91	2.07
Adjacent complementary angles	4.00	.00	3.08	2.23
Section 3				
Isosceles triangle vertex base	3.57	.79	2.67	1.75
Isosceles triangle base vertex***	4.00	1.00	1.70	.68
Isosceles triangle base base	3.14	.38	3.40	1.27
Triangle sum***	3.43	.79	1.40	.97
Section 4				
Equilateral triangle**	3.71	1.25	1.71	.95
Triangle exterior angle	2.50	.55	2.75	.96
Triangle remote interior angle	3.00	.00	2.50	1.00
Section 5				
Corresponding angles	3.00	.00	3.33	2.52
Alternate exterior angles	2.75	.50	3.67	2.52
Alternate interior angles	2.00	.00	4.67	1.16
Supplementary interior angles	4.00	.00	5.00	1.00

*** significant at $P = .001$ level, ** significant at $P = .01$ level, * significant at $P = .05$ level

Discussion

The results of the lab study indicate that the adaptive fading condition outperformed the two non-adaptive conditions (problem solving and fixed fading) on both the immediate and delayed posttest. Additionally, the adaptive fading condition needed significantly fewer worked steps than the fixed fading condition, which indicates that overall the students' knowledge levels increased faster in the adaptive condition. Lastly, no differences were found on immediate and delayed posttest performance between the fixed fading condition and the problem-solving condition.

These findings partially confirm the hypotheses and emphasize the need of the experiment being replicated in the actual classroom to see if the results hold in an environment which contains inherently more noise (e.g., distractions such as announcements over the intercom, students arriving late, off-task behavior, absenteeism, informal peer helping in ways that may or may not be effective etc.).

Experiment 2: Classroom study

Method

Participants

The study took place at a vocational school in a rural area near Pittsburgh where the Geometry Cognitive Tutor is used as a part of the regular geometry instruction. The participants consisted of three 9th grade classes with 51 students led by one teacher. In order to assign the students to the conditions, the student list was sorted based on the students' prior grade in the course. The first three students were then randomly assigned to one of the three conditions, followed by the second three students on the list, and so on.

It should be noted that the vocational school where this study was conducted is a career and technology center where students from surrounding high schools take classes. On average, the students attending such vocational schools tend to have difficulty learning mathematics. Furthermore, since the students have to travel from their respective high school to the vocational school for their classes, there was a substantial level of absenteeism in this study. Therefore, this classroom study provided a challenge as to whether the findings from the lab study could also hold up in this "unfavorable" real-life environment.

Materials

Overall, the materials were very similar to the German lab study with a few differences. First, since the students were already familiar with the Cognitive Tutor, we did not provide instruction up front about how to use the tutor. Instead, the teacher explained to the students what the differences were between the standard Cognitive Tutor and the two example-enhanced versions in technical terms. Second, in contrast to the lab study the Cognitive Tutor's mastery learning mechanism was used during this classroom study in all three conditions due to the agreement with the participating school that our study would not be detrimental to student learning. As such it was of vital importance to give students the possibility to master all skills. As a consequence the tutor presented students with remedial problems for the theorems they had not yet fully mastered until all theorems were mastered (according to the tutor's estimate of the student's mastery level described above). As a result, different students completed slightly different sets of problems. Third, the in vivo study covered more material and had a longer duration than the lab study.

The study comprised all five sections in the tutor curriculum that deal with the geometric properties of angles, including the unit that was used in the lab study. New problems were developed for all units, as our fading procedure required problems that involve particular skill combinations. More specifically, similar to the lab study, the problems in all five sections were sequenced from simpler to more complex; with one-step problems presented first, followed by two-step problems, and eventually by three-step problems. Figure 3 shows a three-step problem in which the enlarged work area shows the worked-out steps and the self-explanation to be provided by the learner. The enlarged table lists the steps in the problem; it is filled automatically by the tutor as the student successfully completes problem-solving steps, and thus shows an overview of the solution at the end of the problem. Furthermore, in four out of the five sections, four new skills were introduced, and in one section, three new skills were introduced, leading to a total number of 19 skills (see Table 2).

Online pretest and posttest presented students with problems covering the same angles theorems as they learned in the Cognitive Tutor. These tests were created with the Cognitive Tutors Authoring Tools (CTAT, see Aleven et al., in press). The pretest and immediate posttest contained the same ten problems, eight of which asked students to indicate whether a step was solvable, and if it was, to provide (1) the value, (2) the theorem used to find the value, and (3) the geometric objects (in the diagram) to which the theorem was applied. The remaining two problems were conceptual knowledge items in which students received a diagram and given measures for a small number of angles. For each of the given angle measures, the students were asked to state which other angle measures could be derived in a single step (i.e., by applying a single geometry theorem).

In addition to the immediate posttest, a delayed posttest was administered. It consisted of six new problems, four of which were procedural items and two were conceptual knowledge items. Since the angles unit is a part of their regular curriculum, participants were not paid for their participation.

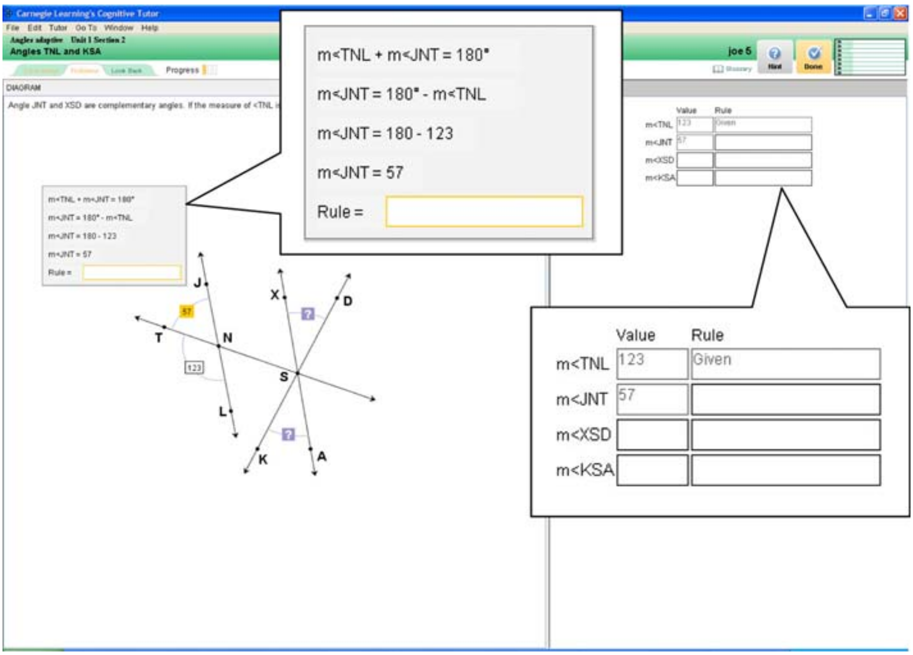

Fig. 3 Multi-step problem with worked-out step in the Geometry Cognitive Tutor

Procedure

First, students took the online pretest and subsequently worked with the Cognitive Tutor for 2 h per week, over a period of 3 weeks, each student in the respective Cognitive Tutor version according to the experimental condition s/he was assigned to. After they finished working on the Cognitive Tutor, students took the online posttest, and 3 weeks later the delayed online posttest was administered.

Results

Posttest performance

Considerable attrition occurred throughout the study, which explains the varying degrees of freedom in the analyses. Of the 51 students ($N = 17$ for each condition) only 20 completed all three tests. Furthermore, 23 students completed both the pretest and the immediate posttest. To test whether an overall pre- to posttest learning gain occurred we compared the pretest scores with both the immediate and delayed posttest scores ($N = 28$). All 28 students completed the pretest of which 23 students completed the immediate posttest and the remaining 5 students completed the delayed posttest. In the analysis of the delayed posttest scores, we included those students who completed at least one other test ($N = 35$) in addition to the delayed posttest. More specifically, of these 35 students who completed the delayed posttest, 20 students completed all three tests; 5 completed the pretest and delayed posttest; and 10 completed the immediate posttest and delayed posttest.

No overall effect ($t < 1$, $N = 23$) was found for students who completed both the pre- and immediate posttests. However, for students who completed both pretest and one of the two posttests ($N = 28$), significant learning occurred in all conditions from pretest ($M = 15.46$, SD = 14.01) to overall posttest ($M = 22.93$, SD = 16.64; $t(27) = 2.27$, $P < .05$, $d = .87$). Additionally, based on Jonckheere-Terpstra tests (one-tailed), no differences between conditions were found on the pretest and immediate posttest (Zs < 1). Lastly, a Missing Value Analysis ($N = 43$) did not yield any significant differences on the immediate posttest ($F < 1$). It is possible that the low compliancy situation of a classroom study in a rural school does not suit the statistical Missing Value Analysis as well as a lab study with a higher compliancy.

Furthermore, using the Jonckheere–Terpstra test (one-tailed), a significant difference ($Z = 1.82$, $P < .05$) between conditions was found on the delayed posttest. The patterns of the means seem to indicate that the effect is mainly due to the superior performance of the adaptive condition. When contrasting the fixed fading condition ($M = 9.38$, SD = 5.25) with the problem solving condition ($M = 8.08$, SD = 4.68), no significant difference in delayed posttest performance was found ($F < 1$). However, when contrasting the adaptive fading condition ($M = 12.80$, SD = 5.61) with the other two conditions (problem solving: $M = 8.08$, SD = 4.68; fixed fading: $M = 9.38$, SD = 5.25) on the delayed posttest, a significant effect in the expected direction was revealed, $t(32) = 2.10$, $P < .05$, $d = .74$. Furthermore, the adaptive fading condition attained a higher performance level on the delayed posttest than the problem solving condition, $t(20) = 2.15$, $P < .05$, $d = .91$. There were no differences in time spent on either of the posttests (Fs < 1).

Number of examples

In order to test whether the adaptive fading condition yielded a different total number of worked-out steps than the fixed fading condition, t tests were conducted on the overall number of examples, and separately on the number of examples for each of the five sections in the Angles unit. First, a t test on the average number of examples over all five sections showed a trend indicating that the adaptive fading condition ($M = 3.65$, SD $= 1.58$) received marginally more examples on average than the fixed fading condition ($M = 2.97$, SD $= .21$), $t(16.59) = -1.77$, $P = .09$, $d = -.61$. Note that the degrees of freedom in this analysis differ from the following analyses since it is based on the total number of worked-out steps over all five sections.

The analysis conducted on the total number of examples in the first section, which was also used in the lab study, showed that the adaptive fading condition ($M = 16.24$, SD $= 7.53$) received significantly more worked steps than the fixed fading condition ($M = 11.65$, SD $= .61$), $t(16.21) = -2.50$, $P < .05$, $d = -.86$. However, in the third section, the adaptive fading condition ($M = 8.10$, SD $= 3.21$) received significantly fewer worked steps than the fixed fading condition ($M = 14.14$, SD $= 2.04$), $t(15) = 4.38$, $P < .01$, $d = 2.25$. Also, in section 4, the adaptive fading condition ($M = 4.71$, SD $= 3.68$) involved marginally fewer examples than the fixed fading condition ($M = 8.43$, SD $= 3.31$), $t(12) = 1.98$, $P = .07$, $d = 1.06$. Note that the degrees of freedom differ between these analyses because the number of students who progressed through the sections decreased with each subsequent section.

When looking at the 19 specific skills across all five sections (the means and standard deviations are provided in Table 2), the following effects were found on five of these skills. In Section 2, the adaptive fading condition required more examples for *linear pair* ($t(16.58) = -2.82$, $P < .05$, $d = -.97$) and *separate complementary angles* ($t(16) = -2.73$, $P < .05$, $d = -.94$). However, in Section 3, the adaptive fading condition involved less examples for *isosceles triangle base vertex* ($t(9.77) = 5.30$, $P < .0001$, $d = 2.70$) and *triangle sum* ($t(14.55) = 4.76$, $P < .0001$, $d = 2.30$). Finally, in Section 5, the adaptive fading condition required fewer examples for *equilateral triangle* ($t(11.19) = 3.36$, $P < .01$, $d = 1.80$). No differences in Cognitive Tutor time, total number of problems solved, and total number of steps (Fs < 1) were found.

Discussion

Despite the considerable attrition in the in vivo study, the results indicate that the adaptive fading condition outperformed the two non-adaptive conditions (problem solving and fixed fading) on the delayed posttest. Additionally, the adaptive fading condition needed marginally more examples than the fixed fading condition, which possibly indicates that overall the students' knowledge levels increased slower in the adaptive condition. Lastly, no differences were found on immediate and delayed posttest performance between the fixed fading condition and the problem solving condition.

These findings partially replicate the findings of the German lab study. While the lab study's finding on the immediate posttest was not replicated, the superior performance of the adaptive fading condition over the other two conditions on the delayed posttest was replicated in the in vivo study. As such, despite the classroom environment containing inherently more noise (e.g., students arriving late, off-task behavior, distractions such as announcements over the intercom, absenteeism, informal peer helping in ways that may or may not be effective etc.) the adaptive fading condition was shown to be a robust instructional method leading to the highest student learning.

Number of examples in section 1 for both studies

Since the lab study included only the first of the five curricular sections covered in the in vivo study, we can compare the effects of the adaptive fading procedures on the number of examples across both studies. The analysis indicates that the in vivo adaptive fading condition ($M = 16.24$, SD $= 7.53$) received significantly more examples than the adaptive fading condition in the lab study ($M = 8.42$, SD $= 4.1$), $t(34) = 3.93$, $P < .001$, $d = -1.29$. This large difference might be explained by the implementation of the experiments. Whereas the students in the lab study all received a fixed number of problems/examples, the students in the in vivo study received a variable (and typically, larger) number of problems/examples because the tutor's mastery learning mechanism was turned on. After the initial sequence of problems (which was the same as the problem set in the lab study), the tutor kept assigning "remedial" problems until the student reached the mastery threshold (meaning that its estimate of the probability of mastering the skill is above .95). If the student made many errors related to a given theorem/skill during these extra problems, the tutor's estimate of mastery of that skill might drop below the lower fading threshold, and the student would be presented with more worked-out examples. However, even when the remedial problems were excluded from the analysis, the in vivo adaptive fading condition ($M = 13.71$, SD $= 5.19$) received significantly more examples than the lab adaptive fading condition, $t(34) = 3.41$, $P < .01$, $d = -1.13$. This difference may reflect differences in the student populations across the two studies. As mentioned, the participants in the in vivo study were vocational students, whereas the participants in the lab study were regular (German) high-school students.

General discussion

Two experiments were conducted comparing "standard" tutored problem solving with a Cognitive Tutor versus two conditions in which tutored problem solving was enriched with worked-out examples. The worked-out examples were faded either in a fixed manner or adaptively based on the quality of the students' self-explanations of worked-out steps. These manipulations were tested both in a lab study and in an actual classroom setting of a regular vocational school. The contrast results of the lab study show that adaptively fading worked-out examples leads to higher performance scores on the immediate *and* delayed posttests than tutored problem solving and fixed fading of examples. The classroom study replicated this effect on the delayed posttest but not on the immediate posttest.

A likely explanation for the fact that the lab results were replicated only partially in the classroom (i.e., no effect of adaptive fading on an immediate posttest) is the use of the Cognitive Tutor's mastery learning criterion in the classroom setting. As mentioned, students in Experiment 2, but not in Experiment 1, received remedial problems for the theorems they had not mastered fully yet (on an individual basis, after they completed the problem sequence described in Table 1). These remedial problems represent additional learning opportunities for students. The mastery learning mechanism may have reduced the group differences in the students' knowledge level (upon completion of the tutor work) in the classroom setting compared to the students in the lab experiment who did not receive any remedial problems. The results indicate further that a possible equalizing effect due to mastery learning did wear off over time, since the adaptive fading condition attained higher delayed posttest performance than the tutored problem solving condition. In other words, even with mastery learning on, a benefit of worked examples is seen.

It is also possible that the diverging results are due partly to the larger amount of "noise" that inherently exists within a real-life environment, as compared to the laboratory. Also, the classroom study took place over a longer period of time with a fairly high level of attrition of students. A considerable number of students missed one (some even missed two) of the three online tests that were given. Yet despite the general difficulty of replicating lab results in the classroom, the current study still shows a benefit of the adaptive fading condition over the standard Cognitive Tutor, a condition which itself has been shown to improve upon classroom instruction (Koedinger et al. 1997, 2000).

Analyzing the number of worked steps students received in the curricular section that was shared across the two studies, we found that the students in the classroom adaptive fading condition received significantly more examples than the lab adaptive fading condition. This difference likely reflects the difference between the student populations in the respective experiments (vocational students vs. high-school students). It may also reflect the many differences between lab and classroom settings. Yet despite this disparity in student populations and environment, the adaptive fading condition showed higher performance outcomes over the tutored problem solving and fixed fading conditions on the delayed posttest.

In contrast to the findings in our previous lab studies (Schwonke et al. 2007), in the current two studies we found no advantage for the fixed fading condition over the problem solving condition, both in terms of students' learning outcomes and in terms of total time spent in the Cognitive Tutor. In our prior lab studies, the fixed fading condition required less time than the problem solving condition. The fact that there was no time difference in the current classroom study (Experiment 2) could be explained by the fact that any efficiency gains due to examples may have been washed out by the tutor's mastery learning mechanism. However, the results from the lab study (Experiment 1) call into question whether any such efficiency gains actually existed in either experiment.

Another possible explanation might be due to the fact that the second of our two previous studies (Schwonke et al. 2007) used think-aloud protocols. Having students think aloud may have contributed to higher transfer performance in terms of conceptual knowledge for the fixed fading condition. It is plausible that the thinking aloud 'forced' students to explicate their strategies of understanding and solving the task at hand which subsequently might have led to an advantage on test items measuring transfer, and that this effect was stronger when students were thinking aloud about examples rather than problems to be solved. This explanation is supported by the fact that in we observed a difference in learning outcomes (and specifically, in students' conceptual understanding) only in the second of the two prior lab experiments, the one during which students were asked to think aloud.

These contrasting findings provide support for the notion that it is important to take differences between the student populations into account when designing example-fading procedures. In our two experiments, we used the same fixed fading procedure, which we developed using data from previous US classroom experiments. While there is much to be said for keeping the fading procedure fixed across experiments, in retrospect, the number of worked steps was an overestimation for the German students in the lab experiment, the number of worked step was an underestimation for the US student in the classroom experiment. An implication of these findings is that future efforts need to carefully determine the appropriate number of worked steps in a tutored problem-solving environment for a specific student population.

Returning to the open question raised in the introduction, the current findings confirm that an adaptive fading procedure based on the quality of students' self-explanations can be effective. The adaptive fading procedure supplied two different student populations with their respective required number of examples, and in both studies, the adaptive fading condition outperformed the combined non-adaptive conditions (problem solving and fixed fading conditions) on the delayed posttests. Additionally, in the lab study, the adaptive fading condition also outperformed the non-adaptive conditions on the immediate posttest.

While our adaptive fading procedure uses self-explanations, the Rapid Dynamic Assessment method (Kalyuga and Sweller 2004, 2005) uses problem solving to guide the fading of worked examples. Despite this difference our results replicate their findings (namely, that adaptively fading examples can be an effective instructional method), but under better controlled conditions where no additional training opportunities that seem to be inherent in the Rapid Dynamic Assessment method, existed in the experimental conditions. Because control conditions in Kalyuga and Sweller (2004, 2005) lacked the additional training opportunities of the experimental condition, it could not be ruled out that the adaptivity effect was (in part) due to more practice. However, both approaches have their merits and have been shown to be effective in the corresponding programs in which they were tested. It would be an interesting research question whether self-explanations and problem solving can be used in complementary ways to optimize the fading to the individual student's increasing knowledge level.

In contrast to the Rapid Dynamic Assessment method (Kalyuga and Sweller 2004, 2005) the Cognitive Tutors used in our study did not take the students' initial knowledge levels into account. They did however, assess the student's subsequent increasing knowledge levels through their performance on explaining example steps. Following the Rapid Dynamic Assessment method's rationale, the current adaptive fading procedure could possibly be further optimized by using the individual student's prior knowledge to determine the initial level of support (i.e., examples). Incorporating this information into the Bayesian knowledge-tracing algorithm that is the basis for adaptive fading method is easy, since this algorithm uses parameters for the expected initial knowledge level of each skill. In the current tutor, that parameter does not vary by student, but it could be set based on a brief assessment prior to each section of the tutor curriculum.

We are also exploring a second way of improving the adaptive fading method. This method is aimed at improving the accuracy of the tutor's estimates of skill mastery determined through knowledge tracing. First, we will set the knowledge-tracing parameters separately for each skill, by fitting them to student–tutor interaction data. This in itself should lead to more accurate estimates of students' mastery levels (Cen et al. 2007; Corbett and Anderson 1995) and may improve adaptive fading of worked-out steps. Second, we will incorporate a new method in the knowledge-tracing algorithm that estimates the probability of student guessing or slipping (Baker et al. 2008), which should also lead to better estimates of student mastery.

In short, the reported findings confirm and extend the findings of Schwonke et al. (2007), which indicated that tutored problem solving combined with fixed fading of worked-out steps leads to more effective learning and better transfer performance. Additionally, the current findings indicate that the implementation of an adaptive fading procedure of worked-out examples within a Cognitive Tutor can be useful in both lab and actual classroom environments across different student populations. As a result from our work Carnegie Learning Inc. (the Cognitive Tutor company) is moving in the direction of

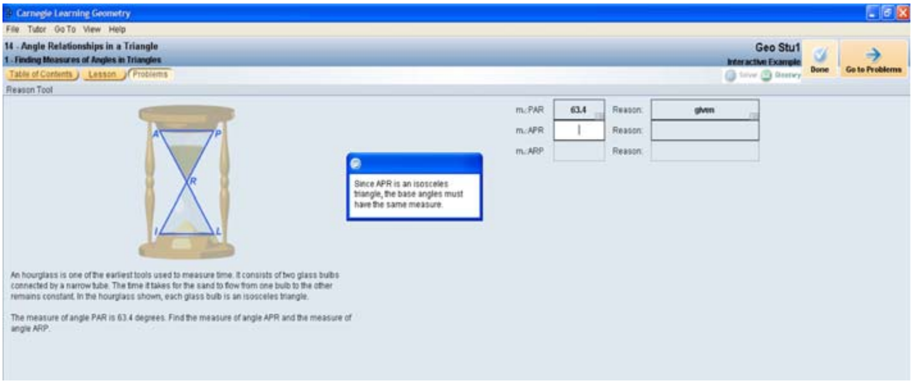

Fig. 4 An interactive example used in the Carnegie Learning Inc. 2008–2009 Cognitive Tutor Geometry™ curriculum

increasing use of worked examples in their 2008–2009 curriculum (see Fig. 4), demonstrating that interactive worked examples, as adjuncts to tutored problem solving, have made the transition into real-world practice. Lastly, these results are in line with the work done by Kalyuga and Sweller (2004, 2005), who showed that adaptive instruction in terms of rapid dynamic assessment during training improved students' learning. Combining their problem-solving assessment with our self-explanations assessment would be an interesting issue to investigate, as well as incorporating levels of student prior knowledge in the self-explanations-based adaptive fading procedure. Future studies that address these issues might find even better ways of matching instructional support with changing levels of student knowledge.

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., & 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., McLaren, B. M., Sewall, J., & Koedinger, K. R. (in press). Example-tracing tutors: A new paradigm for intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*.

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

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.

Baker, R. S. J. d., Corbett, A. T., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. In B. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the 9th international conference on intelligent tutoring systems* (pp. 406–415). Berlin: Springer Verlag.

Cen, H., Koedinger, K. R., & Junker, B. (2007). Is over practice necessary?—improving learning efficiency with the Cognitive Tutor through educational data mining. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proceedings of 13th international conference on artificial intelligence in education* (*AIED2007*) (pp. 511–518). Amsterdam: IOS Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*, 253–278.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-entailed instruction. *Educational Psychology Review, 19*, 509–539.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558–568.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development, 53*, 83–93.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review, 19*, 239–264.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30–43.

Koedinger, K. R., Corbett, A. T., Ritter, S., & Shapiro, L. (2000). Carnegie learning's cognitive tutor$^{\text{TM}}$: Summary research results. (Available from Carnegie Learning, Inc., Pittsburgh, PA): http://www.carnegielearning.com.

McLaren, B. M., Lim, S., Yaron, D., & Koedinger, K. R. (2007). Can a polite intelligent tutoring system lead to improved learning outside of the lab? In R. Luckin, K. R. Koedinger, & J. Greer (Eds), *Proceedings of the 13th international conference on artificial intelligence in education (AIED-07), Artificial Intelligence in Education: Building technology rich learning contexts that work* (pp. 433–440). IOS Press.

Renkl, A., & Atkinson, R. K. (2007). An example order for cognitive skill acquisition. In F. E. Ritter, J. Nerb, E. Lehtinen, & T. O'Shea (Eds.), *In order to learn: How the sequence of topics influences learning* (pp. 95–105). New York, NY: Oxford University Press.

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

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education, 70*, 293–315.

Roy, M., & Chi, M. T. H. (2005). Self-explanation in a multi-media context. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 271–286). Cambridge: Cambridge Press.

Schwonke, R., Wittwer, J., Aleven, V., Salden, R. J. C. M., 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.

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.

Van Lehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, A., Taylor, D., Weinstein, A., & Wintersgill, M. (2005). The Andes physics tutoring project: Five years of evaluations. *International Journal of Artificial Intelligence in Education, 15*, 1–47.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7*, 1–39.