ELSEVIER

Cognitive Science 26 (2002) 147–179

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor

Vincent A.W.M.M. Aleven*, Kenneth R. Koedinger

Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Received 5 February 2001; received in revised form 14 January 2002; accepted 17 January 2002

Abstract

Recent studies have shown that self-explanation is an effective metacognitive strategy, but how can it be leveraged to improve students' learning in actual classrooms? How do instructional treatments that emphasizes self-explanation affect students' learning, as compared to other instructional treatments? We investigated whether self-explanation can be scaffolded effectively in a classroom environment using a Cognitive Tutor, which is intelligent instructional software that supports guided learning by doing. In two classroom experiments, we found that students who explained their steps during problem-solving practice with a Cognitive Tutor learned with greater understanding compared to students who did not explain steps. The explainers better explained their solutions steps and were more successful on transfer problems. We interpret these results as follows: By engaging in explanation, students acquired better-integrated visual and verbal declarative knowledge and acquired less shallow procedural knowledge. The research demonstrates that the benefits of self-explanation can be achieved in a relatively simple computer-based approach that scales well for classroom use.

Keywords: Self-explanation; Metacognition; Computer-based learning environments; Intelligent tutoring systems; Cognitive tutors; Problem solving; Learning; Instruction; Implicit and explicit learning

1. Introduction

A problem for many forms of instruction is that students often come away with shallow knowledge. The students may learn just enough to pass a test, but they lack a deeper understanding of the subject

* Corresponding author. Tel.: +1-412-268-5475; fax: +1-412-268-1266.
E-mail address: aleven@cs.cmu.edu (V.A.W.M.M. Aleven).

0364-0213/02/$ – see front matter
PII: S0364-0213(02)00061-7

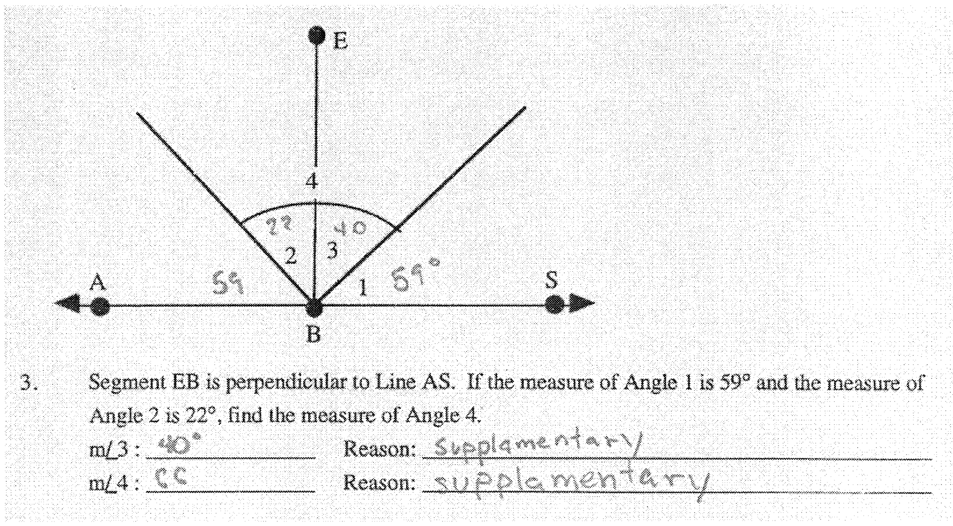

Fig. 1. Example of a student’s shallow reasoning.

matter and have difficulty applying what they have learned to unfamiliar situations. Such lack of transfer is often referred to as “inert knowledge” (Bereiter & Scardamalia, 1985; Cognition and Technology Group at Vanderbilt, 1990). Shallow learning occurs in many forms of instruction and in many domains. In physics, for example, there are many common misconceptions, such as confusing mass and weight or the direction of velocity and acceleration. Novices classify physics problems by superficial features that are not necessarily related to the solution method; experts classify problems based on the type of solution method (Chi, Feltovich & Glaser, 1981).

As another example, novices learning a programming language may learn shallow tactics to select from the built-in functions that the language offers. For example, students learning Lisp may learn the shallow rule: “when dealing with lists, select ‘append’ as the constructor function” (Corbett & Trask, 2000). This rule may work quite often, but it is wrongly-contextualized, since it ignores the desired structure of the list to be constructed.

In geometry, students often rely on superficial visual features, such as the fact that angles look the same in the diagram, even if they cannot be shown to be so by reasoning logically from theorems and definitions. Such heuristics are often successful, but are likely to fail in more complex problems, as is illustrated in Fig. 1. Asked to find the unknown quantities in the diagram, a student makes the inference that the angle on the left has the same measure as angle 1 on the right, which is 59°. These angles look the same, but this inference is not justified.

A long-standing goal of educational research is to help students to avoid shallow learning and instead, to help them learn with understanding, as evidenced by improved transfer (Judd, 1908; Katona, 1940; Simon, 1987; Singley & Anderson, 1989; Wertheimer, 1959), retention, removal of misconceptions (Chi, 2000), and ability to explain acquired knowledge in their own terms. But how to achieve this in actual classrooms?

Many researchers have focused on metacognitive processes that facilitate knowledge construction as a way to get students to learn with greater understanding (Flavell, 1979;

Palincsar & Brown, 1984; Schoenfeld, 1987). This line of research has yielded very interesting instructional programs that elaborate, make visible, support, and help students reflect upon metacognitive processes that are conducive to the construction of knowledge. A number of these programs have been demonstrated to be very effective in actual classrooms (Palincsar & Brown, 1984; Brown & Campione, 1996; Cognition and Technology Group at Vanderbilt, 1990, 1996; Scardamalia, Bereiter & Lamon, 1996; Schwartz, Yerushalmy & Wilson, 1993; White, Shimoda & Fredericksen, 1999).

However, the step from classroom research to widespread use remains a difficult one, in particular for instructional programs that require teachers to substantially alter their approach to teaching. For example, in the domain of geometry, "[c]omputer technology is a powerful support for teaching through guided inquiry, but this approach still depends on teachers, who often find it extremely difficult to carry out in classrooms" (Wiske & Houde, 1993, p. 212). Certainly there are notable exceptions, such as the communities of learners of Brown and Campione (1996) or the anchored instruction developed by the Cognition and Technology Group at Vanderbilt, 1996), but only a small portion of instructional programs whose effectiveness has been demonstrated in classroom research is in regular use in a large number of schools.

In the current research, we share the hypothesis that a focus on metacognition is key to getting students to learn with greater understanding. We focus on a particular metacognitive strategy, self-explanation. A number of studies have shown that students learn better when they explain instructional materials to themselves (Bielaczyc, Pirolli & Brown, 1995; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Ferguson-Hessler & De Jong 1990; Renkl, 1997), or when they explain their own problem solving steps (Ahlum-Heath & DiVesta, 1986; Berardi-Coletta, Buyer, Dominowsky & Rellinger, 1995; Berry, 1983; Gagne & Smith, 1962). Also, there is some evidence that students learn more when they give explanations than when they receive explanations (Brown & Kane, 1988; Webb, 1989), depending on the quality of their explanations (Lovett, 1992).

While these studies suggest that self-explanation could be leveraged to improve educational practice (Renkl, Stark, Gruber & Mandl, 1998), it is still an open question how instructional methods that scaffold or emphasize self-explanation compare to instructional treatments that do not emphasize self-explanation. None of the studies mentioned above compared a self-explanation condition against other (proven) instructional treatments. A related open question is what kind of instruction or scaffolding is needed and most effective in supporting learning through self-explanation in classrooms. The literature indicates that learning through self-explanation is not easy to do. Not all students self-explain spontaneously. There are considerable individual differences in students' ability to self-explain (Chi et al., 1989; Renkl, 1997). Some studies indicate that few students are good self-explainers (Renkl et al., 1998). The challenge therefore is to support students who are not inclined to self-explain or are not good at generating effective self-explanations, without making great demands of an instructor's time.

Some approaches to scaffolding self-explanation that have been shown to work in the laboratory are not likely to meet this challenge. First, effective self-explanations can be elicited through prompting (Chi, de Leeuw, Chiu & Lavancher, 1994; Renkl et al., 1998). But prompting is not feasible in a classroom context, since a single teacher could not possibly provide timely prompts for each individual student. Further, prompting does not benefit all students as much as would be ideal (Renkl et al., 1998, p.106).

Second, self-explanation can be facilitated through an instructional program that involves significant portions of (human) one-on-one instruction (Bielaczyc et al., 1995). But it is not clear that this program could easily be implemented in a classroom context. Even if it could, it seems likely that considerably more continued guidance is needed for the average high school student than the subjects in this study, who were students at one of America's elite universities. Nonetheless, such a program might be a valuable complement to other approaches.

The current study looks at how computer-based instruction can help in scaffolding self-explanation. In two classroom experiments, we evaluate the added value of support for self-explanation in the context of an innovative high-school geometry curriculum called Cognitive Tutor Geometry™. This curriculum, which was developed by our research group, involves about equal parts of classroom instruction and problem-solving practice with intelligent instructional software, a "Cognitive Tutor." Cognitive Tutors provide individualized support for guided learning by doing (Anderson, Corbett, Koedinger & Pelletier, 1995), to an extent that would be difficult to achieve in a traditional classroom setting, where a single teacher has very little time to spend with each individual student. Cognitive Tutors make it possible for a teacher to spend more time helping those students who need help the most. In addition to a geometry tutor, our research group has developed Cognitive Tutors for introductory computer programming and high school Algebra I and Algebra II. A number of studies have shown that Cognitive Tutors help raise students' mathematics achievement relative to traditional mathematics courses (Anderson et al., 1995; Koedinger, Anderson, Hadley & Mark 1997; Koedinger, Corbett, Ritter & Shapiro, 2000). A third-party evaluation study indicates that Cognitive Tutors lead to an increase in student motivation and measurably transform classroom culture (Schofield, 1995). At the time of this writing, the Cognitive Tutor software and curriculum for Algebra I are being used in almost 700 schools across the United States.[1] The Geometry Cognitive Tutor course is in use nationwide in about 100 schools. Thus, these curricula have gone a long way in making the transition from classroom research to widespread use (Corbett, Koedinger & Hadley, 2001), offering an example of how basic cognitive science findings can guide the development of effective instructional technology.

In spite of this success, there is room for improvement: First, while Cognitive Tutors have been shown to be more effective than classroom instruction and more effective than *normal* human one-on-one tutors, they are not yet as effective as the *best* human tutors. Normal human tutors, who have modest domain knowledge and limited training and experience in tutoring, are about 0.4 standard deviations better than traditional classroom instruction (Cohen, Kulik & Kulik, 1982; Graesser, Person & Magliano, 1995). The effect of accomplished tutors is unknown but researchers estimate that the effect size is about 2.0 standard deviations, compared to classroom instruction (Bloom, 1984; Kulik & Kulik, 1991). Cognitive Tutors have been shown to be 1 standard deviation better than classroom instruction (Anderson et al., 1995; Koedinger et al., 1997) and have been estimated to be 1.75 standard deviations better than self-study (Corbett, 2001) and therefore fall in between the two extremes.

Second, like most other forms of instruction, Cognitive Tutors are not immune to the shallow learning problem, as we found out during a formative evaluation of the Geometry Cognitive Tutor. In a classroom experiment involving 71 students in two schools, we evaluated an early version of the tutor, which provided guidance for problem solving but did not provide support for self-explanation (Aleven, Koedinger, Sinclair & Snyder, 1998). We found that there were significant learning gains, due to the combination of classroom instruction and problem-solving practice with the tutor. We also found that the students involved in the study were better at finding answers to test problems than they were at articulating the reasons that were presumably involved in finding these answers, the geometry theorems and definitions. This discrepancy may be due in part to the fact that it is difficult to articulate visual knowledge. But at least in part this discrepancy seems to indicate that students acquired shallow heuristics, such as "if angles look the same in the diagram, they are the same," as illustrated in Fig. 1. Thus, the practical motivation for the study reported in this paper was the desire to improve an existing and successful Cognitive Tutor, by adding facilities for scaffolding self-explanation.

An important research question is how best to support self-explanation? In what format should the system let students state their explanations? What kinds of help and feedback should the system offer to students as they try to produce explanations? These choices are not mere implementation details but are likely to affect students' learning outcomes. The cognitive science literature provides some guidance but not enough to make reliable predictions about learning.

In the current work, we investigate the effectiveness of a relatively simple format for stating explanations: Students explain their problem-solving steps by selecting from a menu the name of the problem-solving principle that justifies the step. This format is interesting because it is easy to understand for students and easy to implement in a computer program.

Further, we take the point of view that the system must provide assistance in the form of feedback on students' explanations and hints on how to explain. Hints are important to help students to proceed when they cannot on their own. Feedback on explanations is important because it may help students to construct better explanations than they would without feedback. This viewpoint was taken also by Conati and VanLehn (2000) in their work on the SE-COACH, a computer tutor to support self-explanation. This viewpoint is also consistent with a number of cognitive science studies on the effect of feedback on explanations (Needham & Begg, 1991; Lovett, 1992). Further, especially when dealing with instructional technology, feedback is important because without it, will students bother to provide explanations at all? They may not feel compelled to respond to prompts for explanations by a computer tutor in the same way that they would respond to a human tutor (Lepper, Woolverton, Humme & Gurtner, 1993; du Boulay, Luckin & del Soldato, 1999). An initial study with a tutor version that provided prompts for explanations but did not provide feedback on explanations confirmed this concern (Aleven, Popescu & Koedinger, 2000). Students provided few explanations and even fewer good explanations.

We are not the only researchers who investigate the effectiveness of computer-based instructional environments designed to support self-explanation, but we are not aware of any studies that compare the instructional effectiveness of such environments to instructional methods that do not emphasize self-explanation. Conati and VanLehn developed the SE-

COACH, an intelligent tutoring system that helps students explain worked-out examples in the domain of physics (Conati & VanLehn, 2000). Renkl (in press) looked at the effect of a computer-based environment that supplements self-explanation with "instructional explanations" that are displayed on the student's request. Both systems were evaluated empirically, but the focus was on comparing different levels of assistance for self-explanation—Renkl but not Conati and VanLehn found a significant effect of the support for self-explanation—not on assessing the added value of self-explanation support over more typical instruction.

In addition to the theoretical issue raised above of when and why to tutor self-explanation, this study also addresses theoretical issues regarding the nature of knowledge acquired through self-explanation. We ask what greater understanding due to scaffolded self-explanation (assuming for the moment that it occurs) means in terms of the underlying knowledge and learning processes. Previous accounts of learning through self-explanation have focused on how self-explanation leads to the construction of declarative knowledge, such as inference rules for applying physics principles or mental models of the circulation system (Chi, 2000). The main mechanisms are gap identification and gap filling (VanLehn, Jones & Chi, 1992). We subscribe to VanLehn et al.'s account, but we note that in the domain of geometry, and in other mathematical and nonmathematical domains as well, expertise involves visual and verbal components and learning processes. An interesting question therefore is how tutored self-explanation might affect such a hybrid learning process.

Based on evidence from a number of studies, Case and colleagues have argued that mathematical understanding develops as students create internal conceptual structures that integrate visual/analog intuitions with discrete formal ways of knowing (Griffin, Case & Siegler, 1994; Kalchman, Moss & Case, 2001). More generally, the integrated acquisition of visual knowledge and formal language/symbols appears consistent with key characteristics of expertise in many domains. Knowledge of sophisticated visual patterns or "perceptual chunks" has been identified as a hallmark of expertise in many domains including chess (De Groot, 1966; Chase & Simon, 1973), electrical circuits, and geometry. Along with this visual knowledge, experts often acquire special-purpose language or symbolic forms, like "knight fork," "voltage," "$V = IR$," "linear pair," or "$\angle ABC \cong \angle EFG$."

In related prior research, Koedinger and Anderson (1990) presented a cognitive model of geometry expertise in which conceptual structures called "diagram configuration schemas" link visual pattern knowledge with (verbal) formal knowledge of geometric properties and constraints. They demonstrated the power of diagram configuration schemas, both empirically, in showing how they explain human experts' leaps of inference in proof planning, and computationally, in using them in a computer model that can efficiently find proofs of geometry theorems. Further, Van Hiele argued that geometry students' knowledge develops in five levels progressing from initial knowledge of visual patterns, through properties, to proof knowledge (Burger & Shaughnessy, 1986).

What was not demonstrated in this past work is how instruction or self-explanation might help students integrate visual perceptual and verbal declarative knowledge and enhance learning. Students' experienced-based learning tends to yield perceptual pattern knowledge acquired through analogy to examples (Anderson & Lebière, 1998). Students' explicit verbal learning yields verbal declarative structures that need to be interpreted to be useful in problem solving—otherwise they remain inert (Bereiter & Scardamalia, 1985; Cognition and

Technology Group at Vanderbilt, 1990) able to be recalled, but not able to support problem solving. One reason that self-explanation may improve learning is that it may help students connect potentially inert verbal knowledge (e.g., geometry theorems) with perceptual patterns used in problem solving. To explore this hypothesis, we created a mathematical model that illustrates the role of various categories of knowledge (shallow, procedural, integrated visual and verbal declarative) on post-test performance. The model also illustrates that self-explanation instruction may yield qualitatively different kinds of knowledge than problem-solving practice.

In the remainder of the paper, we first describe the Geometry Cognitive Tutor. We then present results from two experiments designed to evaluate the added value of self-explanation in a classroom. In both experiments, we compared instruction with two different versions of the Geometry Cognitive Tutor, one that supports self-explanation in the context of problem solving, one that supports problem solving without self-explanation. Finally, we present our mathematical analysis aimed at finding out more about how self-explanation instruction affects the nature of acquired knowledge.

2. Supporting self-explanation in a Cognitive Tutor: learning by doing and explaining

In this section, we describe the Geometry Cognitive Tutor that was used in our experiments. Cognitive Tutors, a kind of intelligent tutoring systems (Wenger, 1987), provide support for *guided learning by doing.* They assign problems to students on an individual basis, monitor students' solution steps, provide context-sensitive feedback and hints, and implement a mastery learning criterion (Anderson et al., 1995).

Cognitive Tutors are grounded in the ACT-R theory of cognition and learning (Anderson & Lebière, 1998). ACT-R, like other theories of cognition, distinguishes between declarative and procedural knowledge. Procedural knowledge is goal-oriented performance knowledge that can be executed efficiently. It is acquired through practice. Within the ACT-R framework, procedural knowledge is modeled as production rules. Declarative knowledge, on the other hand, cannot be executed directly, but can be applied to specific goals by general interpretive procedures. Declarative knowledge is inert without such procedures (Bereiter & Scardamalia, 1985; Cognition and Technology Group at Vanderbilt, 1990). In ACT-R, declarative knowledge is modeled as "chunks." It includes verbal as well as visual knowledge and is acquired more or less directly through perception, instruction, or reading. Procedural knowledge is *implicit knowledge* that is not available to awareness whereas declarative knowledge is *explicit knowledge* that we are aware of in visual or verbal form (c.f., Dienes & Perner, 1999).

Cognitive Tutors take from ACT-R the notion that the skills involved in a complex task can be decomposed and modeled as production rules. Each Cognitive Tutor employs a cognitive model, expressed as a set of production rules, that represents the skills and strategies of students at various levels of competence. The cognitive model enables the tutor to perform the task that the students are working on, as is typical of intelligent tutoring systems (Wenger, 1987). The tutor uses the model to analyze an individual student's

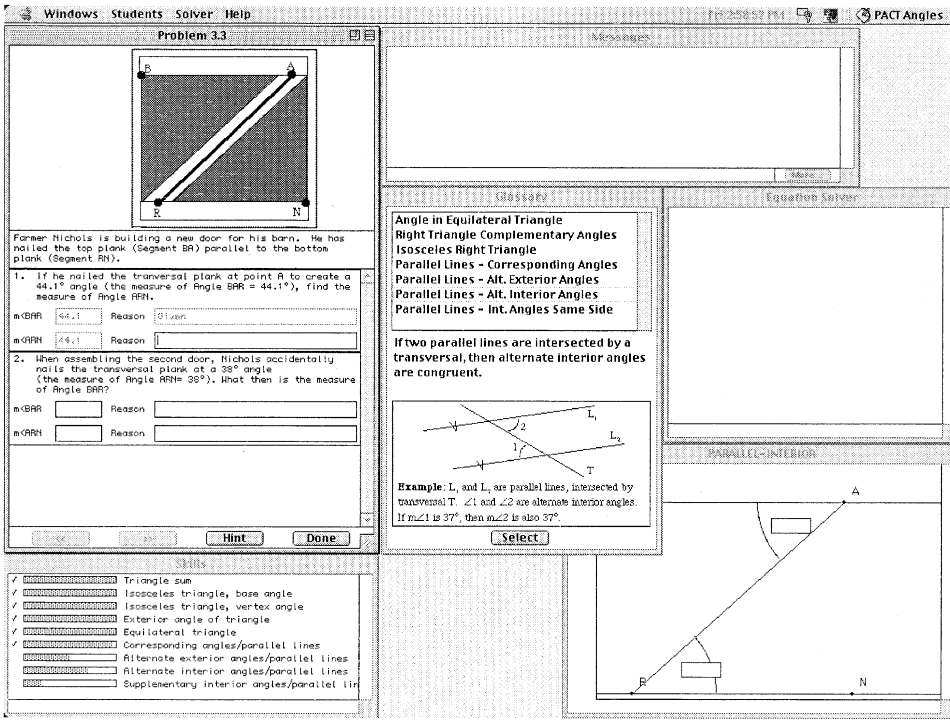

Fig. 2. The new version of the Geometry Cognitive Tutor, with support for self-explanation of solution steps.

problem-solving performance in terms of the underlying production rules ("model tracing"). The tutor uses that analysis to update its student model, which records the probability that the student masters each crucial production rule ("knowledge tracing") (Corbett & Anderson, 1995).

The Geometry Cognitive Tutor, shown in Fig. 2, is an integrated part of a full-year high-school geometry course, which was developed by our research group following guidelines by the National Council of Teachers of Mathematics (NCTM, 1989). In this course, students spend about 40% of classroom time solving problems on the tutor. The remaining time is devoted to lectures, classroom discussion, and small-group activities. The tutor curriculum consists of units dealing with area, the Pythagorean theorem, angles, similar triangles, circles, and quadrilaterals. The Angles unit (the subject of our study) consists of problems in which the student is given a diagram and is asked to compute unknown quantities such as angle measures or the area of a triangle. Many problems involve a real-world context, as illustrated in the problem displayed in the top left window shown in Fig. 2, which is an early problem dealing with the alternate interior angles theorem.

In order to study the added value of self-explanation during problem-solving practice, we created a new version of the Geometry Cognitive Tutor, adding facilities to support self-explanation. The new version supports *guided learning by doing and explaining:* The tutor requires that students enter correct solutions to geometry problems and that they explain all steps correctly. Students can enter explanations in a straightforward manner, by typing the name of the problem-solving principle that justifies the step ("explanation by reference"). For example, a student could explain a step in which the triangle sum theorem was applied by typing "Triangle Sum." In order to facilitate the process of providing of explanations, the tutor provides a Glossary of geometry knowledge, shown in the middle of Fig. 2. The Glossary lists relevant theorems and definitions, illustrated with short examples. It is meant to be a reference source which students can use freely to help them solve problems. Students can enter explanations by selecting a reference from the Glossary, as a convenient alternative to typing the explanation.

The Geometry Cognitive Tutor provides feedback on the students' solutions as well as their explanations. It displays error messages in response to the most common errors. Further, it provides on-demand hints, with multiple levels of hints available for each step, as is detailed further below. The tutor keeps track of the student's mastery of each skill to be learned. It displays its estimates in a skillmeter window, shown on the bottom left in Fig. 2. In order to complete a section of the tutor curriculum, students need to bring all skills above the mastery level threshold, which is set at 95%.

In the example shown in Fig. 2, which is based on an actual log file of a student working with the tutor, the student explained a problem-solving step, using the Glossary to figure out which geometry rule justified the step. He conducted a focused Glossary search, apparently based on a cue extracted from the problem, for example, the use of the term "parallel" in the problem statement. He looked at three of the four rules dealing with parallel lines listed in the Glossary, in the middle of Fig. 2. The Glossary shows a statement of each rule, illustrated with an example diagram. He then selected the right reason, "alternate interior angles." Interestingly, the student consulted the Glossary in spite of the fact that he got the answer right on the first attempt. This suggests that he may have found the answer by applying a shallow heuristic of the kind we discussed before. By studying the relevant rule in the Glossary, the student may have picked up on verbal cues in the rule and may have used them to improve and better integrate his visual and verbal geometry knowledge (into a diagram configuration schema as described in Koedinger & Anderson, 1990). For example, the student may have better connected the phrase "alternative interior angles" with a visual Z-shaped image. This example underscores the potential value of having students explain steps by providing references to problem-solving principles.

If the student had not been able to complete this explanation step by himself, he could have asked the tutor for a hint. The tutor's hints were designed to communicate a general strategy for knowledge search: If you do not know something, use an available resource, such as the tutor's Glossary, to look it up—incidentally, the student in the example discussed above carried out this strategy on his own. For most steps, multiple levels of hints are available, as illustrated in Table 1, which shows the hints for the next step in Fig. 2, namely, to explain why the measure of angle ARN is 44.1 degrees. The initial hints suggest that students search the Glossary for an applicable geometry rule. More detailed hints state an applicable rule and summarize how it is applicable. The hints for numeric answer steps (as opposed to explanation steps) follow the same general plan but go into more detail about how an applicable rule can be applied to find an unknown quantity. The student manages the hint levels. For any given step, help starts at level one. The hint level is increased by one for each subsequent help request.

Table 1
Example hint sequence generated by the geometry cognitive tutor, annotated in the rightmost column to show the underlying hint plan developed and used by the designers

	Hint text	Underlying hint plan
1.	You gave the correct answer. The measure of Angle ARN is 44.1 degrees. If you had to make an argument to explain your answer, what reason would you give?	State the goal.
2.	Two parallel lines (Segment BA and Segment RN) are cut by a transversal (Segment AR). How did you use this fact to find the measure of Angle ARN?	Suggest cue to focus search for applicable geometry
3.	Some rules dealing with parallel lines are highlighted in the Glossary. Which of these reasons is appropriate? You can click on each reason in the Glossary to find out more.	Suggest that student search the Glossary and use cue to narrow down the search.
4.	When two parallel lines are intersected by a transversal, alternate interior angles are equal in measure. That is why the measure of Angle ARN is equal to the measure of Angle BAR.	State applicable rule and summarize how it applies to the problem at hand.
5.	Enter "alternate interior angles" as the reason.	Say what to type.

The Geometry Cognitive Tutor is different from another intelligent tutoring system that supports self-explanation, the SE-COACH developed by Conati and VanLehn (2000). In the Geometry Cognitive Tutor students explain their own solution steps and do so by naming the problem-solving principles that justify the step. In the SE-COACH, students explain the steps in worked-out examples and construct explanations of problem-solving principles using a structured interface with menus and templates. It is not clear *a priori* how each choice affects students' learning.

3. Experiment 1: evaluation of a Cognitive Tutor that supports self-explanation

The goal of Experiment 1 (as well as that of Experiment 2) was to test the hypothesis that problem-solving practice with a Cognitive Tutor leads to deeper learning if the Cognitive Tutor requires that students explain their steps. We compared the learning results of students working with two versions of the Geometry Cognitive Tutor, a version that supported self-explanation in the manner described above and a version that provides no support for self-explanation. Students in both conditions also received classroom instruction, as detailed below.

3.1. Participants

Experiment 1 took place in a suburban high school near Pittsburgh. It involved 41 students taking the Cognitive Tutor Geometry course, two periods taught by the same teacher and his assistant. The students ware mostly 10th graders, that is, 15 and 16-year olds. The students were assigned to the Explanation condition or the Problem-Solving condition on the basis of their prior scores in the same course, in such a way that the conditions were balanced. The prior scores were based on a number of quizzes, tests, and homework assignments, all taken prior to the experiment. The prior scores do not involve the pretest. Of the 41 students who started the experiment, 24 students completed it, 13 students in the Problem-Solving condition, and 11 students in the Explanation condition. Some students did not complete the experiment because they ran out of time at the end of the semester, other students because the teacher forgot to assign a post-test.

3.2. Pretest and post-test

The pretest and post-test were designed to assess students' ability to solve geometry (angles) problems and to assess their understanding of the relevant geometry theorems and definitions. We created six test forms and assigned them randomly to students to counterbalance for test difficulty. The same forms were used for the pretest and the post-test. Each test form included six problems for a total of 30–33 test items per form. The tests included regular items as well as transfer items. The regular test items were much like the steps that the students solved while working on the tutor. That is, in regular test items, the students were asked to compute an unknown quantity in a given problem diagram and had to provide a reason why their answer was correct, in terms of geometric theorems. The students were provided with a "reason sheet" listing the relevant geometry rules and were told that they could freely reference it. Subsequently, we refer to the steps in these problems as Answer items and Reason items.

The criterion for grading the explanations was whether students were able to justify their answers in terms of geometry definitions and theorems. An explanation was considered to be correct when the student gave the name of the right geometry theorem or gave a correct statement of the right theorem, possibly in their own words. An explanation was considered to be incorrect for example when all that was said was that two particular angles were congruent ("$\angle 1$ is equal to $\angle 2$"), or when the student reiterated in words the arithmetic that had been carried out to find the numeric answer (e.g., "add 62 and 41 together"), or when the explanation left out some of the elements of a geometry theorem (e.g., "all the angles have to equal 180" or "because L1 is parallel to segment L2").

Three of the problems in each test included transfer items. In these problems, as in the other problems, students were presented with a problem statement and a diagram. However, instead of being asked to compute the value of certain unknown quantities, the students were asked to judge whether there was enough information to find the values of unknown quantities. If in their judgment there was enough information to uniquely determine the value of such a quantity, they were asked further to compute the value and state an explanation. Otherwise, they could simply answer "No" meaning that there was not enough information to determine a unique value. Items within these problems that could be computed, based on available information, were grouped with the Answer and Reason items. Items for which no answer could be found due to missing information are subsequently referred to as "Not Enough Info" items. During their work on the tutor curriculum, students had not encountered items of this type. Therefore, these items provide a measure of how well skills learned with the tutor transfer to unfamiliar nut related problems. Each test form included 12–14 Answer items, 12–14 Reason items, and 5–7 Not Enough Information items.

3.3. Procedure

The experiment was conducted in the context of the regular geometry instruction and curriculum. The instruction focused on one unit of the tutor curriculum, the tutor's Angles unit, which deals with the geometric properties of angles. Each day of the week, students had one class period of classroom activities and one class period of working problems on the computer, using the Geometry Cognitive Tutor. All students took the pretest shortly before they started working on the Angles unit. The students completed the post-test shortly after finishing the tutor work. Since the work on the tutor is to a large degree self-paced, different students started working on the Angles tutor unit at different times, and took different amounts of time to complete the tutor unit. In order to complete the tutor unit, the students had to satisfy the tutor's mastery level criterion. That is, (1) they had to complete the required problems for each of the three sections of the Angles unit, (2) they had to complete any remedial problems that the tutor might select, based on its student model, until they reached mastery for all targeted skills in the Angles unit. Thus, due to the fact that the tutor selects remedial problems on an individual basis, different students solved different sets of problems. Some of the classroom instruction was given before students took the pretest, some of it was given in between the pretest and post-test.

Students in the two conditions worked with different tutor versions. Students in the Explanation condition were required to explain problem-solving steps. They used the Geometry Cognitive Tutor as described above and shown in Fig. 2. The students in the Problem-Solving condition were not required to explain their problem-solving steps; they had to provide only the correct solution steps. They used a version of the Geometry Cognitive Tutor that was different in two ways from that used by the Explanation condition students: There was no column for entering reasons in the tutor's answer sheet. Also, the tutor's hints did not suggest that students use the Glossary in order to find applicable geometry knowledge, but only stated which rule could be applied and how. Other than this, there were no differences between the two conditions.

3.4. Results

We ran a $2 \times 2 \times 3$ mixed ANOVA on students' test scores, with condition as a between subjects factor and test time (pretest v. post-test test) and item type (Answer, Reason, and Not Enough Information) as within subject factors. There was a main effect of test time ($F(1,22) = 20.96$, $p < .0001$), indicating that students' test scores increased from pretest to post-test (see Fig. 3). Further, there was an interaction between condition and test time ($F(1,22) = 10.3$, $p < .005$), indicating that students in the Explanation Condition improved significantly more than their counterparts in the Problem-Solving Condition.

There was no 2-way interaction between condition and item type and no 3-way interaction between condition, test-time and item type. Therefore, students in the Explanation improved uniformly more than students in the Problem Solving condition on all three types of items (see Fig. 3).

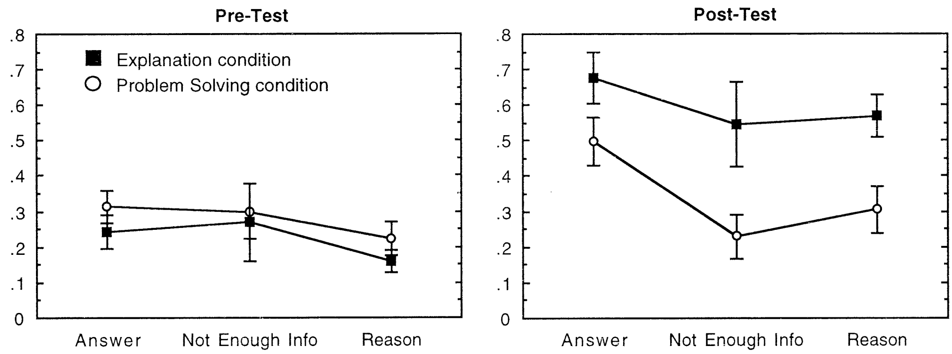

Fig. 3. Test scores in Experiment 1.

Students in the Explanation condition spent about 18% more time working on the tutor than did students in the Problem-Solving condition (436 ± 169 minutes for the Explanation condition, 368 ± 160 for the Problem-Solving condition). The difference is not statistically significant ($F(1,22) = 0.1.01$, $p = .33$). The main reason for this difference is that students in the Explanation condition had more work to do per problem, since they had to explain their solution steps. On the other hand, students in the Explanation condition needed fewer problems to fulfill the tutor's mastery level criterion: 102 ± 40 v. 136 ± 53 problems. The difference was marginally statistically significant $F(1,22) = 3.06$, $p < .1$). The mastery level was the same for both conditions.

Of the students who completed the experiment, the Explanation group had slightly better prior course grades (87.3 vs. 83.3). That difference is not statistically significant ($F(1,22) = 1.38$, $p = .25$). As mentioned, at the outset both conditions were balanced in terms of their prior grades in the course, due to the way students were assigned to conditions.

3.5. Discussion

The results suggest that there are considerable learning benefits to having students explain their steps during problem solving practice with a Cognitive Tutor. Not surprisingly, it leads to better performance in providing reasons for solution steps. Also, training on reason-giving transfers to better overall performance in both providing answers and making judgments about whether there is enough information to give an answer. This strongly suggests that self-explanation during guided learning by doing leads to greater understanding. Interestingly, these performance differences were obtained even though the students who explained their answers during training solved fewer problems. In other words, they did fewer problems, but got more out of each problem. On the other hand, we cannot entirely rule out an alternative interpretation, namely, that the students in the Explanation condition performed better because they spent slightly more time on the tutor or were a slightly better than average sample (although this is not evident in the pretest scores). While it seems unlikely that this explanation would account for all post-test and learning gain differences between the conditions, concern about these issues motivated us to do another experiment.

4. Experiment 2: evaluation of a Cognitive Tutor that supports self-explanation, controlling time on task

The goal Experiment 2 was again to test the hypothesis that students learn with greater understanding when they explain their solution steps, as they work with a Cognitive Tutor. The purpose was to replicate the results of Experiment 1, while controlling for time on task. As in Experiment 1, we compared an Explanation condition and a Problem-Solving condition, working with different tutor versions. To make sure that both groups spent the same amount of time on the tutor, we changed the criterion for finishing the tutor work: Instead of a mastery level criterion, as in Experiment 1, we used a time limit, as is explained below.

4.1. Participants

The experiment took place in the same suburban school as Experiment 1 and involved 53 students in two periods of a geometry course. Ten subjects were excluded from the analysis because they did not complete all activities. Two further subjects were excluded, one who left 75% of the post-test blank, one who spent less time on the tutor than required.

Students were assigned to an Explanation condition and a Problem-Solving condition. Unlike in Experiment 1, we assigned the students in each period to a separate condition. This was done so that students in different conditions would not work in the computer lab at the same time. In Experiment 1, students in the Explanation condition had sometimes complained that they had more work to do than students in the Problem-Solving condition, because they had to give reasons for their solution steps. 19 students in the Explanation condition and 22 students in the Problem-Solving condition completed the experiment.

4.2. Pretest and post-test

The same pretest and post-test were assigned as in Experiment 1.

4.3. Procedure

All students carried out the same four activities as in Experiment 1. They worked through the tutor's Angles unit, received classroom instruction related to this unit, and took a pretest and post-test. The only difference between the two experiments was the criterion for completing the tutor. In the current experiment the tutor used a time limit of 7 hours. This was the average amount of time spent by students in the Explanation condition in Experiment 1. To make sure that students made reasonable progress through the assigned tutor unit, we also included time limits for each of the three sections that make up this unit. Students were advanced out of the first two sections when they exceeded the time limit for the given section, 2:20 hours and 2:30 hours respectively, or reached mastery level for all skills of the section, whichever came first. They completed the whole unit when they had spent 7 hours total.

As before, the tutor assigned problems for students to work on. Within each section, it first

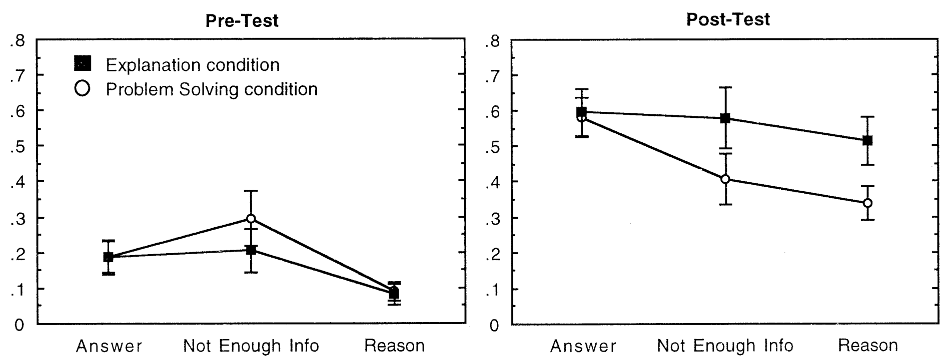

Fig. 4. Test scores in Experiment 2.

assigned a fixed sequence of required problems, followed by remedial problems based on the tutor's assessment of the student's skills, as captured in the student model.

4.4. Results

Both conditions spent an equal amount of time on the tutor (Explanation group: 511 ± 73 min, Problem Solving group: 503 ± 63). These numbers are higher than the tutor's time limit of 7 hours because they include students' idle time, whereas the tutor factored out idle time as it kept track of the amount of time spent by each student. Students in the Explanation group completed half the number of problems completed by their counterparts in the Problem Solving group: 83 ± 24 versus 166 ± 66 ($F(1,39) = 27.2$, $p < .0001$).

We ran a 2 × 2 × 3 ANOVA on the test scores, with condition as between-subjects factor and test time (pretest v. post-test) and item type (Answer, Reason, and Not Enough Info) as within-subject factors. There was a main effect of test time ($F(1, 39) = 69.4$, $p = .0001$). Thus, there were performance gains from pretest to post-test, just as we had found in Experiment 1 (see Fig. 4). There was a marginally significant interaction between test time and condition, ($F(1, 39) = 3.83$, $p < .06$), suggesting that students in the Explanation condition learned more than students in the Problem-Solving condition (see Fig. 4). Finally, there was a 2 × 2 × 3 interaction ($F(2, 78) = 3.15$, $p < .05$).

In order to understand the 2 × 2 × 3 interaction, we conducted two orthogonal planned comparisons. We used the students' learning gains as dependent measure, as this is our primary criterion for the effectiveness of the instruction. The gain is the relative improvement over the pretest score. It is defined as:

$$\frac{\text{post-test score} - \text{pre-test-score}}{1 - \text{pre-test score}}$$

A priori, we expected that the advantage of the Explanation condition would be greater on items that hinge on deeper understanding (i.e., the Reason and Not Enough Info items) than on items where shallow knowledge can have some degree of success (i.e., the Answer items). We did not have any expectations as to which of the two types of items requiring deeper understanding would show the greatest difference in learning gains.

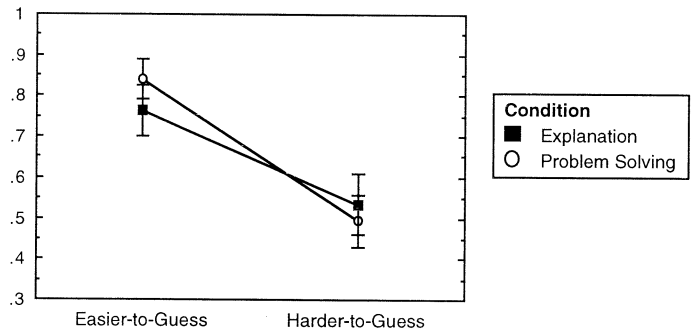

Fig. 5. Post-test scores in Experiment 2 (proportion correct), for Easier-to-Guess and Harder-to-Guess Answer items.

As expected, the first planned comparison found a difference between Answer items on the one hand and Reason and Not Enough Info items grouped together on the other hand ($F(1, 37) = 7.76$, $p < .01$). Thus, the conditions' gain scores differed more on Reason and Not Enough Info items than on Answer items, with the Explanation group having the greater gain scores (see Fig. 4). The second comparison found no difference between Reason items and Not Enough Info items ($F(1, 37) = 0.65$, $p > .4$). In other words, it was not clear that the conditions' gain scores were wider apart on the one item type than on the other. In short, the orthogonal planned comparisons show that students in the Explanation condition learned better than students in the Problem-Solving condition to deal with the items requiring deeper understanding, the Reason items and Not Enough Info items. The fact that students in the Explanation condition learned better to explain their steps is important in its own right. The guidelines from the National Council of Teachers of Mathematics emphasize mathematical communication as an important objective for mathematics education (NCTM, 1989).

In order to further investigate issues of deep learning, we divided the post-test Answer items into Easier-to-Guess items and Harder-to-Guess items. We defined these categories as follows: An unknown quantity in a problem is Easier-to-Guess when it is equal to a quantity from which it can be derived in a single step. Otherwise, a quantity sought is Harder-to-Guess. Guessing heuristics such as "if angles look the same in the diagram, then their measures are the same" are likely to be successful on Easier-to-Guess items despite the lack of understanding they reflect, but not on Harder-to-Guess items.

As shown in Fig. 5, the Explanation condition students performed better on the Harder-to-Guess items while the Problem-Solving students performed better on the Easier-to-Guess items. The difference however does not reach statistical significance ($F(1,39) = 1.72$, $p < .20$). Further, we calculated for each student the proportion of her total number of errors that were errors of commission, as opposed to errors of omission (see Fig. 6). We ran a 2×2 ANOVA on the proportion of commission errors, with condition as a between-subjects factor and "guessability" (Easier-to-Guess v. Harder-to-Guess) as a within-subject factor. We found a main effect of guessability ($F(1, 16) = 5.94$, $p < .05$) and a marginally significant effect of condition ($F(1, 16) = 4.44$, $p < .06$). Students in the Explanation condition tended to make more errors of omission, students in the Problem-Solving condition tended to make more errors of commission. Together, these two analyses suggest that students in the

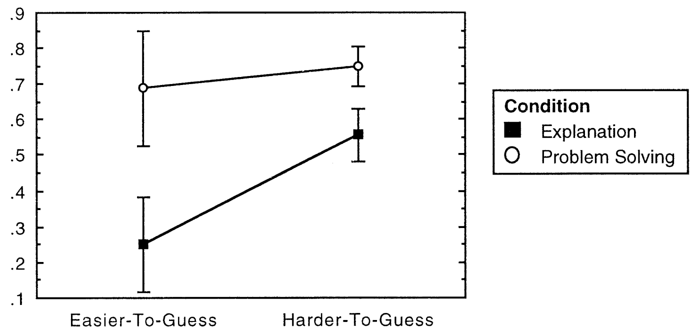

Fig. 6. Errors of commission on the post-test (Experiment 2), as proportion of the total number of errors. Errors were classified either as errors of commission or errors of omission.

Explanation condition were more likely to succeed on harder problems and more likely to reflect on the sufficiency of their knowledge, rather than guess or jump to incorrect conclusions.

5. Comparison of training data between the conditions

In order to study in more detail why the students in the Explanation condition learned with greater understanding, we analyzed the data from the log files of students' actions with the tutor. We did this analysis for Experiment 2 only.

First, we were interested in comparing the learning rate between the conditions. Having established that self-explanation helps students to learn with greater understanding, we were interested to see whether there was evidence that self-explanation helps students to learn problem-solving skills faster (i.e., with less practice). Second, we were interested in the data on students' Glossary use. *A priori*, Glossary use seemed likely to affect students' declarative knowledge of domain principles. Therefore, any observed differences between the conditions in students' Glossary use might help to explain the different learning outcomes.

5.1. Comparison of learning rates

The post-test results suggest that self-explanation speeds up the learning of problem-solving skills. Students in the Explanation condition did slightly better on the Answer items of the post-test than students in the Problem Solving condition, even though they had solved only half the number of problems during training. Therefore, we expected that the training data would show faster learning of problem-solving skills by the Explanation group, in the form of a faster increase in the success rate.

As an aggregate measure of the learning rate, we compared the training performance on *all* steps by students in the Explanation condition, namely, 237 steps, against performance on the same number of steps by students in the Problem-Solving condition. (A step is a subgoal in a tutor problem, that is, a numeric answer or explanation to be entered into the tutor's answer sheet. A step was considered to be correct if the student made no errors and took no

Table 2
Students' performance with the tutor in experiment 2 (percentage correct)

	Explanation condition		Problem-Solving condition	
	Success rate	Number of steps	Success rate	Number of steps
Numeric Answer steps	51%	237	56%	357
First 237 steps	51%	237	51%	237
Rest of the steps	N/A		62%	220
Explanation steps	55%	236	N/A	

hints.) To this end we compiled a comparison set, which included 52% of all steps by students in the Problem-Solving condition, for a total of 237 per student on average. We included in this comparison set, for each student in the Problem-Solving condition, the *earlier* steps for each skill, leaving out the later steps. The steps were selected in such a way that the proportion of steps by a given student involving a given skill was the same in the comparison set as it was in the complete set of all steps by students in the Problem-Solving condition. The steps were mapped to skills following the tutor's cognitive model.

Using this measure for the learning rate, we found no difference between the groups (see Table 2). The data indicate that both conditions performed equally well during the first 237 problem-solving steps (51% correct). The Problem Solving condition then went on and improved, achieving a success rate of 62% on the rest of the steps. Thus, the tutor data do not confirm that explanation increases the rate of learning of problem-solving skills.

Of particular interest is the fact that the students in the Problem Solving condition did better on Answer items during training but not on the post-test. This indicates that the students in the Explanation condition achieved better transfer of skills learned with the computer tutor to a different environment, such as a paper and pencil test, further evidence that the students who explained their answers learned with greater understanding.

5.2. Comparison of Glossary use

We were interested to learn whether students' patterns of Glossary use differed between the Problem Solving and Explanation conditions, and if so, how these differences this might help to explain the different post-test results. As mentioned, the tutor's Glossary is a reference source that students could use freely as they were working problems on the tutor. For each relevant geometry rule, the Glossary contained a statement and a short example illustrated with a diagram. In the process of consulting the Glossary, the students might read and interpret descriptions of problem-solving principles, study examples, or reason about the applicability of the problem-solving principles to the problem at hand. These activities would likely enhance their (verbal and visual) declarative knowledge of problem-solving principles. On Numeric Answer steps, we expected that students in both conditions would use the Glossary regularly as they were solving problems. Specifically, we expected that the students would consult the Glossary when they realized (as a result of self-monitoring) that a step was beyond their capabilities. Therefore, we expected that they would not make many errors without consulting the Glossary first. Further, we expected that on Explanation steps the

Table 3
Frequency of Glossary use in experiment 2 (percentage of steps with Glossary use)

	Numeric answer steps		Explanation steps	
	Explanation condition	Problem-Solving condition	Explanation condition	Problem-Solving condition
Glossary used	5.2%	1.5%	47%	N/A
Used deliberately	3.4%	1.4%	15%	

students in the Explanation condition would use the Glossary often in order to select explanations. It seemed likely that this would carry over to their Glossary use on Numeric Answer steps. Therefore, we expected that students in the Explanation condition would use the Glossary more often on numeric answer steps than students in the Problem-Solving condition.

We measured frequency of Glossary use as the percentage of steps for which the students inspected at least one Glossary item.[2] Further, we defined *deliberate* Glossary use as follows: A step involved deliberate Glossary use if the student inspected at least one Glossary item for at least one second.[3] Deliberate use may be a better gauge of effective learning activities than overall Glossary use. As shown in Table 3, the Glossary use on Numeric Answer items was very low in both conditions. The frequency of deliberate use is 3.4% in the Explanation condition and 1.4% in the Problem-Solving condition. Thus, the prediction that students would use the Glossary in an effort to avoid errors was not borne out. Students made many errors without consulting the Glossary. For Numeric Answer steps, the error rate was 49% in the Explanation condition and 44% in the Problem Solving condition. The actual rate of Glossary use is far below this. Thus, on Numeric Answer steps, the students did not use the Glossary as often as seemed appropriate.

Students in the Explanation condition used the Glossary about half the time when they explained their solution steps (47% of Explanation steps). However, the students used the Glossary in a deliberate manner on only 15% of the Explanation steps. Most of the time when they used the Glossary, they also *selected* references from the Glossary (45% of the explanation steps), as opposed to typing the explanation.

One possible explanation for the low Glossary use on Numeric Answer steps is that the given student population lacks the necessary metacognitive and mathematical reading comprehension skills to take advantage of the Glossary. They may not make good judgments about the difficulty of steps and thus not realize when they could benefit from using the Glossary. Further, it may often be too difficult for them to find, interpret, and apply relevant information in the Glossary. In order to get students to learn how to use a resource like the Glossary (a secondary goal we had when we added the Glossary), the tutor must provide more support for doing so (Aleven & Koedinger, 2000a).

In sum, students in both conditions rarely used the Glossary to solve problems, but students in the Explanation condition frequently used it to explain problem-solving steps. Thus, one benefit of the Explanation Condition seems to have been that it encourages the use of descriptions and examples of problem-solving principles in the Glossary.

6. A mathematical model of acquired knowledge constituents

A major finding of this research is that self-explanation does not increase the *rate* at which knowledge is acquired as much as it changes the *nature* of the knowledge acquired. We wanted to better understand this change in the nature of the acquired knowledge. In particular, we wanted to investigate whether the experimental data offer support for the hypothesis that self-explanation during problem solving helps students better to integrate verbal and visual knowledge. To do so, we created a mathematical model of hypothesized internal knowledge constituents and attempted to fit it to the pattern of post-test performance data. The model comprises (1) dependent variables corresponding to four measures of post-test performance, (2) three independent variables, capturing the strength of different knowledge constituents at the post test, and (3) four equations relating the dependent and independent variables, that is, relating the strength of the different knowledge constitutents to post-test perfomance. While quantitative in nature, the model is meant primarily as a tool for thinking qualitatively about the experimental data, as discussed further below.

The model distinguishes between four aspects of post-test performance and contains a variable for each. The aspects are: Score on the Easier-to-Guess Numeric Answer items (variable E), score on Harder-to-Guess Numeric Answer items (H), score on Reason items (R), and score on Not Enough Info items (N). These types of post-test items were described before. The model is based on the assumption that students' post-test performance is a result of a mix of three types of knowledge. For each of these types of knowledge, there is an independent variable in the model that represents the probability that students have acquired this type of knowledge: Correct Procedural Knowledge (p), Shallow Procedural Knowledge (s), and Declarative Knowledge (d).

Declarative Knowledge (d) integrates visual and verbal knowledge of the problem-solving principles, such as the knowledge a student might acquire by active processing of visual images and associated verbal descriptions in the tutor's Glossary or in a classroom discussion. As mentioned, this integrated knowledge can be represented with diagram configuration schemas (Koedinger & Anderson, 1990), which link visual patterns with formal geometric constraints and properties. For example, a diagram configuration schema dealing with isosceles triangles would relate an image of a triangle that looks like an ice cream cone or Christmas tree in any shape or rotation with the properties that the triangle has two congruent sides and that the two angles opposite the congruent sides are congruent. The schema also contains information specifying which subset(s) of these properties are sufficient to infer that the schema applies ("ways-to-prove") and that all of the listed properties hold. For example, to infer that the isosceles triangle schema applies, either one of the properties mentioned above must hold. The other property then follows. Visually recognizing the ice cream cone shape helps to retrieve the schema from memory but is not sufficient to decide that the schema applies—what looks like an isosceles triangle may not always be so.

Procedural knowledge, as defined in the ACT-R theory of cognition (Anderson & Lebière, 1998), is implicit knowledge that can be used to perform a task but is not open to reflection. In our domain, Correct Procedural Knowledge (p) enables a student to find numeric answers to the types of geometry problems presented in the experiment. Shallow Procedural Knowledge (s) is incorrect procedural knowledge, guessing heuristics that are overly general or

Table 4
Justification for the equations in the model

Item type	Equation	Justification To be successful one needs to
Numeric, Easier-to-Guess	$E = d+(1\text{-}d)p+(1\text{-}d)(1\text{-}p)s$	Have declarative knowledge (d), or have correct procedural knowledge (p), or guess right (s)
Numeric, Harder-to-Guess	$H = d+(1\text{-}d)p$	Have declarative knowledge (d) or correct procedural knowledge (p)
No Enough Info	$N = d+(I\text{-}d)(1\text{-}s)$	Have declarative knowledge (d) or refrain from guessing (1-s)
Reason	$R = d+(1\text{-}d)*0.25$	Have declarative knowledge (d) ro narrow down using keywords (0.25)

wrongly contextualized, such as “angles that look the same, are the same” or “if a triangle looks like a Christmas tree standing upright, its two bottom angles are equal,” the latter being a shallow version of the isosceles triangle schema discussed above. We suspect (and the examples just given illustrate) that this knowledge is primarily visual and lacks integrated verbal constraints. In the Christmas tree example, the missing verbal knowledge is that two sides of the triangle are congruent. Without this constraint, the schema applies also to triangles that look like isosceles triangles but are not known to be so.

Finally, the model has a set of equations that capture the ways in which the different types of knowledge can be brought to bear to solve the four different types of test items. These equations, listed in Table 4, observe the ACT-R theory. In short, Declarative Knowledge is useful for all four types of test items, assuming that a student has mastered the necessary procedures for interpreting and applying the declarative knowledge. Correct Procedural Knowledge is useful for finding numeric answers, whether Easier-to-Guess or Harder-to-Guess. Shallow Procedural Knowledge helps in getting the Easier-to-Guess answer items right, but gets in the way when dealing with Not Enough Info items. We describe the motivation for each equation in more detail.

6.1. Numeric, Easier-to-Guess items

In order to find a numeric answer one can use either declarative knowledge (d) of the relevant problem-solving principles plus an appropriate interpretive procedure, or one can use procedural problem-solving knowledge (p). Also, on Easier-to-Guess items Shallow Procedural Knowledge (s) is likely to succeed.

6.2. Numeric, Harder-to-Guess items

Harder-to-Guess items can be solved using (correct) procedural problem-solving knowledge (p) or by interpreting relevant declarative knowledge (d). Shallow procedural knowledge or guessing heuristics (s) however usually do not apply to these types of items, in contrast to the Easier-to-Guess items. Thus, they do not affect performance on the Harder-to-Guess items.

6.3. Not Enough Info items

As mentioned, when dealing with Not Enough Info items, students must recognize that the value of a quantity cannot be uniquely determined, based on the available information (i.e., the quantity is underconstrained). One way to arrive at correct answers to Not Enough Info items is through deliberate processing of a verbal declarative encoding (*d*) of geometric constraints (stored in the ways-to-prove slot of diagram configuration schemas). That is, students must retrieve from memory a small set of diagram configuration schemas that are potentially relevant, cued by either visual or verbal information extracted from the problem. Then they must verify that none of the retrieved schemas apply. For each retrieved schema they must verify that no ways-to-prove are satistied, meaning that there is a missing geometric constraint, (e.g., that certain lines are parallel) or a missing numeric condition (e.g., that certain quantities are known). This way, they may come to a reasoned conclusion that no relevant geometry knowledge applies and that the answer is, "no, not enough information is available." Procedural knowledge cannot be deliberately interpreted in this way. It can only be executed directly (Anderson & Lebière, 1998).

Even without relevant declarative knowledge, students may get Not Enough Info items right, by (1) trying to solve them using their correct procedural knowledge (*p*), (2) failing to find an answer, and (3) concluding that there is not enough information. In order to rely on procedural knowledge in this way, students must not use shallow guessing heuristics (*s*) of the kind that we have seen before, such as the shallow rule "if angles look the same, they are the same." These guessing heuristics might produce a numeric answer to Not Enough Info items, which of course is incorrect.

6.4. Reason items

A student may explain a numeric answer by retrieving from long-term memory a verbal, declarative encoding (*d*) of the relevant domain principle, perhaps checking that it actually applies to the current problem, and then verbalizing that encoding. Procedural problem-solving knowledge cannot be used to explain answers, because it is not available for verbalization (Anderson & Lebière, 1998). Further, given that students were using a reason sheet during the post-test, we assumed there is about a 25% chance that a student might get an explanation right without having accurate Declarative Knowledge. Using keywords or visual cues, it is usually possible for students to narrow down the set of possible reasons on the reason sheet to about four. Thus, we added a fixed parameter to represent this 25% chance that a students gets the reason correct without correct Declarative Knowledge.

When we fit this model to the post-test data of Experiment 2, using the Generalized Reduced Gradient method offered by the Microsoft Excel Solver,[4] we find the values of the independent variables (i.e., those representing the strength of knowledge constituents) shown in Table 5. With respect to the quality of the fit between model and data, the correlation between the actual and predicted values of the post-test variables (shown in Table 6) is $r = 0.995$ when we consider the *average* post-test performance in the two conditions and is $r = 0.49$ when we consider the fit *for each individual student*. While these correlations suggest that a reasonable fit is achieved between model and data, we are not primarily interested in

Table 5
Probability of mastery of the various knowledge types, found by fitting the model variables to the post-test data of experiment 2

Knowledge type	Variable	Explanation condition	Problem-Solving condition
Shallow Procedural Knowledge	*s*	0.58	0.68
Correct Procedural Knowledge	*p*	0.30	0.42
Declarative Knowledge	*d*	0.32	0.12

the statistical properties of the fit, nor in the particular variable estimates that were found. An ideal model would include additional parameters, both to represent student variability and to represent variability in the different production rules associated with different post-test items. Such a model is beyond the scope of the current paper.

However, we do think the current model helps to develop a useful qualitative interpretation of the different pattern of post-test results across the two conditions. The model suggests, plausibly, that the Explanation condition acquired more reliable (integrated visual and verbal) declarative knowledge (higher *d* value), whereas the students in the Problem-Solving condition had stronger procedural knowledge, but also a greater inclination to guess, as evidenced by the greater probability for both Correct and Shallow Procedural Knowledge (higher *p* and *s* values, see Table 5). Thus, students in the Explanation condition were better at explaining their problem-solving steps because they had more integrated verbal and visual declarative knowledge. They were better able to deal with Not Enough Info items because they had stronger declarative knowledge and less shallow procedural knowledge. Students in the Problem Solving condition did better on the Easier-to-Guess items due to stronger procedural knowledge, as well as a greater amount or weighting of shallow knowledge. The students in the Explanation condition were able to do better on the Harder-to-Guess items, using declarative knowledge to make up for their weaker procedural knowledge.

Further insight can be had from a “qualitative sensitivity analysis” of the model, namely, by asking which internal knowledge constituents are needed to explain the statistically significant differences observed between the conditions. The answer follows from the following qualitative argument: The significant difference between conditions on the Reason and Not Enough Info transfer items requires an internal knowledge structure better acquired by the Explainers. The Declarative Knowledge parameter fills this role. Given the Declar-

Table 6
Comparison of the predicted values of the input variables of the model against the post-test data of experiment 2

Type of test item	Variable	Explanation condition		Problem-Solving condition	
		Actual	Predicted	Actual	Predicted
Numeric, Easier-to-Guess	*E*	0.76	0.80	0.84	0.84
Numeric, Harder-to-Guess	*H*	0.54	0.52	0.49	0.49
Not Enough Info	*N*	0.58	0.61	0.41	0.41
Explanation	*R*	0.51	0.49	0.34	0.34

ative Knowledge difference in favor of Explainers and the fact that the conditions did equally well overall on the Answer items (i.e., the Easier-to-Guess and Harder-to-Guess items grouped together), there is a need for a second type of internal knowledge structure that is better acquired by the Problem Solvers. The Correct Procedural Knowledge parameter fills this role. Finally, the observed interaction between the Harder-to-Guess and Easier-to-Guess test items (Explainers did better on the Harder-to-Guess items, but Problem Solvers did better on the Easier-to-Guess items) requires a third internal knowledge structure that raises Problem Solvers' performance only on Easier-To-Guess items. The Shallow Procedural Knowledge parameter fills this role. Thus, this sensitivity analysis reveals that, within the structure of the proposed model, dropping out any one of the parameters would result in an inability to capture a characteristic difference in post-test performance.

We suspect that the students in the Explanation condition also acquired a limited amount of shallow explanation knowledge, for example, knowledge associating names of geometry theorems with particular visual configurations. This knowledge would enable them to do well on the Reason items of the post-test, but would not necessarily represent deeper understanding. However, if this knowledge were the only difference between the conditions, we could not explain why the Explainers did better on Not Enough Info items and Harder-to-Guess Answer items. While explanation "by reference" as supported in the Geometry Cognitive Tutor does not stop all shallow reasoning, it is fair to conclude that it reduces the chances of shallow reasoning, both at the problem-solving and the explanation level.

7. Discussion

We have shown how a simple computer-implemented form of self-explanation support leads to enhanced student learning. We found that student achievement gains acquired from such support are quantitatively better than achievement gains from tutored problem solving. More importantly, we found a qualitatively different pattern of learning and post-test performance across the conditions. This different pattern of learning and performance appears to be well accounted by a model of more shallow procedural acquisition in the problem-solving control condition and more integrated visual and verbal declarative knowledge acquisition in the explanation condition.

7.1. Learning process

How does having students explain their problem-solving steps yield these differences in knowledge acquisition? Our account is similar to the computational model proposed by VanLehn et al. (1992). They stressed the ways in which self-explanation leads students to find and repair gaps in their knowledge and thus construct new declarative knowledge. We share with VanLehn et al. a concern for distinguishing deep and superficial modes of learning. However, our account is different in that it is in the context of the ACT-R theory and in that it emphasizes the role of visual and verbal modes of learning. Many domains, like geometry, have a significant visual component. We believe self-explanation aids learning in such domains because it facilitates the construction of more integrated visual and verbal declarative knowledge. Further, it reduces students' reliance on shallow visual inference and implicit procedural knowledge that is not available for reflective reasoning.

Shallow learning and understanding in geometry problem solving is often a result of (a) starting out with weak declarative knowledge, both verbal and visual, and (b) relying primarily on visual information during problem-solving practice. Without constraints from verbal declarative knowledge, students are subject to the frailties of induction. They are prone to be captured by specific visual patterns and acquire overly general or wrongly contextualized production rules. A primary example in the geometry domain is "if it looks the same, it is the same."

When students engage in an effort to explain their problem-solving steps they strengthen their verbal declarative knowledge and better integrate visual and verbal knowledge. They are more likely to retrieve verbal declarative descriptions of problem-solving principles from long-term memory and thus strengthen those memories. Alternatively, they may consult the Glossary to read the verbal descriptions and view visual examples (as Explainers more often did) and thus strengthen and connect these forms of knowledge. Most importantly, when students attempt to explain they may reflect on implications of both forms of knowledge to root out shallow visual knowledge and incorrect interpretations of verbal forms. Explaining requires more deliberate processing that can lead students to identify and fill knowledge gaps (VanLehn et al., 1992). Processing new terms can cue a student's attention toward critical features of visual instances and so guide more accurate rule induction. For instance, a student may currently have shallow or incomplete knowledge that is roughly, "if isosceles, then equal angles." Processing "*base* angles" during explanation may highlight the need for an additional condition, cue the student's attention to the visual symmetry of the base angles, and so facilitate the induction of a new condition: "if angles at the base of an isosceles . . . ."

In sum, we submit that the deliberate processing of verbal declarative knowledge required by self-explanation complements students' natural inclination toward example-based induction. This combination of induction and verbal learning modes is consistent with studies that show that students learn better from instruction based on both examples and rules than from instruction based on either one alone (Holland, Holyoak, Nisbett & Thagard, 1986). Our emphasis on the importance of the integration of visual and verbal forms of knowledge is consistent with research on memory advantages of a visual-verbal "dual code" (Paivio, 1971). It is also consistent with research on expertise that has identified perceptual chunks (visual knowledge), in addition to more jargon (verbal knowledge), as distinguishing experts from novices (e.g., Chase & Simon, 1973; Koedinger & Anderson, 1990). Finally, our emphasis on visual and verbal knowledge integration is consistent with other successful instructional interventions that have been fundamentally designed to help students connect visual intuitions and more formal verbal knowledge (Griffin et al., 1994; Kalchman et al., 2001).

7.2. Effect of providing feedback on self-explanations

The current study shows that self-explanation can be supported effectively using instructional software which provides feedback on explanations and other forms of scaffolding, such as on-demand hints and a Glossary. In earlier studies, it was shown that self-explanation helped students learn with greater understanding, even if they did not receive feedback on their explanations (e.g., Chi et al., 1989). How important then was the scaffolding provided by the tutor in the current study, in particular the feedback on explanations? Although the data do not enable us conclusively to single out any given tutor feature for credit or blame, there is evidence that the tutor's feedback on students' explanations was very important.

First, the tutor's feedback helped students improve their explanations. As mentioned, when students enter explanations, the tutor tells them if the explanation is correct or not. Further, the tutor insists that students provide a correct explanation for each step. Students in the Explanation condition got explanations right on the first try on 55% of the steps. This suggests that the tutor's feedback and hints contributed to the eventual correctness of as much as 45% of the students' explanations. Further, the tutor's feedback on the students' *numeric answers* may also have helped students in getting the *explanation* right, since knowing the correct numeric answer obviously provides constraints on the explanation of that answer. It is quite likely that without these forms of scaffolding, the success rate on explanations would have been considerably lower, which may well have had a negative impact on students' learning outcomes.

The importance of correctness and quality of self-explanations has been debated in the cognitive science literature (see Conati & VanLehn, 2000). A number of researchers have pointed out that the quality of explanations matters (Lovett, 1992; Renkl et al., 1998). On the other hand, one researcher has argued that even incorrect self-explanations can be beneficial (Chi, 2000), as the resulting incorrect "knowledge" is likely, sooner or later, to trigger further self-explanations and thus be corrected. But what happens if those further self-explanation are again incorrect? Chi's argument seems to depend on a premise that most self-explanations will be correct. And indeed, across her studies, 75% of students' self-explanations were correct (Chi, 2000). But this high level of explanation ability is not likely in all domains. For example, in the current study, the success rate was only 55%. In such domains, it is likely that feedback on explanations is crucial or at least very useful if students are to learn by self-explaining. Further, in many domains, poorer students are likely to benefit from feedback on explanations. A similar argument motivated the design of the SE-COACH system (Conati & VanLehn, 2000).

The tutor's feedback was important in a second way: without tutor feedback students may not have attempted to explain their steps in the first place. Both concerns were confirmed in a study that involved a version of the tutor that prompted students to supply explanations in their own words, but did not analyze the explanations or provide feedback (Aleven & Koedinger, 2000b). Without feedback, students often ignored these prompts and provided very few high-quality explanations on their own.

7.3. Effect of explaining steps by reference

In the current study, students' explanations were references to the problem-solving principles involved. Students could explain steps simply by naming the problem-solving principle that justified the step. In most other studies on self-explanation, by contrast, students explained in their own words. Perhaps the greatest surprise of the current study was that this relatively simple format has a positive effect on learning. Apparently, explanation by reference leads to some amount of conscious and conscientious processing of verbal declarative knowledge and to better integration of visual and verbal knowledge, as discussed above. Tacit reasoning can lead to learning even when it is not verbalized (Berardi-Coletta et al., 1995).

An interesting open question is whether students learn more when they explain in their own words, when working with a computer tutor. This question is especially relevant since natural language processing technology has progressed to the point that researchers are now exploring its use in tutoring systems (Rose & Freedman, 2000). Having students explain in their own words is likely to have a number of benefits. When students explain in their own words, the tutor can require that they provide complete explanations, not just references to problem-solving principles as in the current tutor. This makes it more likely that students attend more fully to the problem-solving principles and that they detect gaps in their own knowledge (cf., VanLehn et al., 1992). Also, when explaining in their own words, it may be easier for students to build on existing knowledge or receive some credit for explanations that are partially correct. The tutor has more information about their thinking processes and thus is in a better position to provide targeted help in the construction of new knowledge, possibly in the form of a dialog (Aleven & Koedinger, 2000b). Finally, when students explain in their own words, there may be a generation effect (Anderson, 1999), due to the fact that one recalls information from memory, rather than recognizing it in a menu. In order to investigate empirically whether and why having students explain in their own words leads to better learning, we are currently adding natural language understanding capabilities to the Geometry Cognitive Tutor (Aleven, Popescu & Koedinger, 2001). The current "explanation by reference" format provides a yardstick against which to measure the effectiveness of this technologically more sophisticated approach.

8. Conclusion

In many forms of instruction, it is difficult to ensure that students avoid the learning of shallow heuristics and that instead, they come away with deeper understanding. Our study shows that self-explanation, supported by means of intelligent instructional software, can help towards this goal and can enhance learning in actual classrooms. We found that *guided* self-explanation adds value to *guided* problem-solving practice without self-explanation. Specifically, we found that problem-solving practice with a Cognitive Tutor is even more effective when the students explain their steps by providing references to problem-solving principles.

Supporting self-explanation in this manner leads to greater understanding, as was manifested in a number of different ways. First, the students who had explained problem-solving steps during training were better at explaining their problem-solving steps and dealt better with transfer problems. Second, while both groups learned equally well to solve geometry problems (i.e., to determine unknown measures in diagrams), it appeared that the students who had explained dealt better with harder-to-guess items, whereas students who did not explain scored higher on easier-to-guess items. Third, the explainers were less likely to jump to unwarranted conclusions (i.e., the proportion of errors of commission was lower). Fourth, the explainers had a greater ability to transfer what was learned from computer-based training to a paper-based post-test. In other words, the knowledge and skills acquired by the nonexplainers appeared to be tied to irrelevant particulars of the tutor environment.

The greater understanding that resulted from self-explanation seems due to qualitatively different knowledge. It can be explained in terms of more integrated visual and verbal declarative knowledge, used more reflectively, and less shallow procedural knowledge. In terms of learning processes, the act of explaining problem-solving steps appears to help students to integrate two modes of learning: implicit visual induction and explicit verbal knowledge acquisition. Example diagrams are easier to interpret than the unfamiliar and abstract terms in verbal rules. However, when students induce general rules based primarily on the visual cues attended to in diagrams, they are likely to form shallow visual heuristics. Bottom-up induction is an error-prone process. When students explain steps, they are forced to pay more attention to the verbal representation of problem-solving principles and verbal cues help focus attention on the critical visual features. Top-down guidance from verbal knowledge helps to avoid the frailties of less reflective induction. An interesting finding in this regard is that the explainers made greater use of the tutor's on-line Glossary of geometry knowledge. This very likely helped them in acquiring stronger and more accurate declarative knowledge that integrates visual patterns and verbal constraints, as in the diagram configuration schemas characteristic of geometry experts (Koedinger & Anderson, 1990).

The current study is one of the first to compare instruction that focuses on self-explanation against a proven instructional condition that does not emphasize or support self-explanation. The control condition in our study was based on a curriculum which involved an earlier version of the Geometry Cognitive Tutor, a version which does not support self-explanation. This curriculum has been shown in earlier studies to be superior to traditional classroom instruction (Koedinger et al., 2000). It is significant that self-explanation can make a difference even when compared to such a high standard.

By showing how self-explanation can be supported by intelligent instructional software, the current study adds to studies reported in the literature that have shown that self-explanations can be elicited through prompting (Chi et al., 1994; Renkl, 1997) and that students can be taught to self-explain (Bielaczyc et al., 1995). Key features of the self-explanation support investigated in this study were that the tutor (1) requires students to explain their problem-solving steps in the form of *references* to problem-solving principles and (2) scaffolds the process of constructing explanations through feedback and hints. Feedback on self-explanations appears to be important, even if our data do not conclusively prove this. Without feedback, students would probably have generated fewer correct explanations, with a detrimental effect on learning. In fact, they may not have had much incentive to provide any explanations at all.

A surprising finding is the fact that self-explanation is effective even when students are asked only to name the problem-solving principles that were involved in each step, but not to state the problem-solving principle or to elaborate how it applies to a problem. Perhaps equally surprising is the fact that self-explanation was scaffolded effectively by a computer tutor. Thus, other aspects of self-explanation that have been hypothesized to be crucial, such as the presence of a human instructor or the fact that students explain in natural language (as opposed to a structured computer interface) are not a necessary condition for effective support. It is an interesting question whether these aspects are even relevant. For example, will students learn more effectively when they are asked to provide more complete explanations or state explanations in their own words? This is a focus of our on-going research.

Our study shows that self-explanation can be leveraged to improve education in actual classrooms. The Geometry Cognitive Tutor with support for self-explanation helps students to deal better with more challenging problems and with transfer items. Further, the fact that this tutor helps students to learn to explain their steps is important in its own right, as communication of mathematical results is an important curricular objective of the National Council of Teachers of Mathematics (NCTM, 1989). Certainly, problem solving and communication are not the only objectives of mathematics education. Also, with respect to supporting curricular objectives such as mathematical connections and communication, more can and ought to be done in classroom discussions than can be supported in today's technology. However, a tough question is how to move teachers in this direction as well as students. We think technology has a role for both. Also, cognitive science studies such as that presented in this paper can help convince and educate teachers as to the value of and techniques for supporting self-explanation.

To conclude, our study illustrates how basic cognitive science research findings can be leveraged to have an impact on educational practice. The Cognitive Tutor for geometry problem-solving, with the self-explanation support described in the current paper, is available commercially and is being used in a growing number of classrooms in the United States.

Notes

1. A company called Carnegie Learning, Inc. has been formed to market the Cognitive Tutors and curricula. See http://www.carnegielearning.com.
2. We include in the count only the steps where students were asked to *infer* a numeric answer, by applying a geometry definition or theorem, or where they were asked to explain such an inference. We excluded those steps where a quantity sought was given in the problem statement.
3. One second is obviously too short to read and interpret a Glossary item, but not too short to recognize an item one has seen before. Also, it may not take much more than one second to rule out the applicability of a Glossary item visually by detecting a mismatch between the problem diagram and the example diagram of the Glossary item.
4. "Microsoft Excel Solver uses the Generalized Reduced Gradient (GRG2) nonlinear optimization code developed by Leon Lasdon, University of Texas at Austin, and Allen Waren, Cleveland State University." From the Excel on-line help.

Acknowledgments

Dr. Bethany Rittle-Johnson provided very helpful comments on an early version of this paper.

The following people contributed to the research reported in this paper: Colleen Sinclair, Karen Cross, Jaclyn Snyder, and Bob Janosko. This research is sponsored by an NSF grant to the Center for Interdisciplinary Research on Constructive Learning (CIRCLE), a research center located at the University of Pittsburgh and Carnegie Mellon University. We gratefully acknowledge their contributions.

References

Ahlum-Heath, M. E., & DiVesta, F. J. (1986). The effect of conscious controlled verbalization of a cognitive strategy on transfer in problem solving. *Memory and Cognition, 14,* 281–285.

Aleven, V., & Koedinger, K. R. (2000a). Limitations of student control: do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 292–303). Berlin: Springer Verlag.

Aleven, V., & Koedinger, K. R. (2000b). The need for tutorial dialog to support self-explanation. In C. P. Rose & R. Freedman (Eds.), *Building Dialogue Systems for Tutorial Applications. Papers from the 2000 AAAI Fall Symposium* (pp. 65–73). Menlo Park, CA: AAAI Press.

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). Lecture Notes in Computer Science 1452. Berlin: Springer Verlag.

Aleven, V., Popescu, O., & Koedinger, K. R. (2001). Towards tutorial dialog to support self-explanation: adding natural language understanding to a cognitive tutor. In J. D. Moore, C. L. Redfield, & W. L. Johnson (Eds.), *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future* (pp. 246–255). Amsterdam, IOS Press

Anderson, J. R. (1999). *Cognitive Psychology and Its Implications* (5th ed.). New York: Worth Publishing.

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

Anderson, J. R., & Lebière, C. (1998). *The Atomic Components of Thought.* Mahwah, NJ: Erlbaum.

Berardi-Coletta, B., Buyer, L. S., Dominowsky, R. L., & Rellinger, E. R. (1995). Metacognition and problem-solving: a process-oriented approach. *Journal of Experimental Psychology: Learning Memory, and Cognition, 21* (1), 205–223.

Bereiter, C., & Scardamalia, M. (1985). Cognitive coping strategies and the problem of "inert knowledge." In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and Learning Skills: Research and Open Question* (vol. 2, pp. 65–80). Hillsdale, NJ: Erlbaum.

Berry, D. C. (1983). Metacognitive experience and transfer of logical reasoning. *Quarterly Journal of Experimental Psychology, 35A,* 39–49.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13,* 221–252.

Bloom, B. S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13,* 3–16.

Brown, A. L., & Campione, J. C. (1996). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice* (pp. 229–270). Cambridge, MA: The MIT Press.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: learning to learn and learning from example. *Cognitive Psychology, 20,* 493–523.

Burger, W. F., & Shaughnessy, J. M. (1986). Characterizing the Van Hiele levels of development in geometry. *Journal for Research in Mathematics Education, 17,* 31–48.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual Information Processing.* New York: Academic Press

Chi, M. T. H. (2000). Self-explaining expository texts: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in Instructional Psychology,* (pp. 161–237). Mahwah, NJ: Erlbaum.

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.

Chi, M. T. H., de Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Cognition and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19* (6), 2–10.

Cognition and Technology Group at Vanderbilt. (1996). From visual word problems to learning communities: changing conceptions of cognitive research. In K. McGilly (Ed.), *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice* (pp. 157–200). Cambridge, MA: The MIT Press.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Education outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal, 19* (2), 237–248.

Collins, A., & Brown, J. S. (1988). The computer as a tool for learning through reflection. In H. Mandl & A. Lesgold (Eds.), *Learning Issues for Intelligent Tutoring Systems* (pp. 1–18). New York: Springer Verlag

Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: a computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education, 11,* 398–415.

Corbett, A. T. (2001). Cognitive computer tutors: solving the two-sigma problem. In M. Bauer, P. Gmytrasiewicz, & J. Vassileva (Eds.) *User Modeling 2001: Proceedings of the Eighth International Conference, UM 2001* (pp. 137–147). New York: Springer.

Corbett, A. T., Koedinger, K. R., & Hadley, W. S. (2001). Cognitive tutors: from the research classroom to all classrooms. In P. Goodman (Ed.), *Technology Enhanced Learning: Opportunities For Change* (pp. 235–263). Mahway, NJ: Erlbaum.

Corbett, A. T., & Trask, H. (2000). Instructional interventions in computer-based tutoring: differential impact on learning time and accuracy. In *Proceedings of the ACM CHI '2000 Conerence on Human Factors in Computing Systems* (pp. 97–104).

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

De Groot, A. (1966). Perception and memory versus thought: some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem Solving.* NY: Wiley.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22* (5).

Du Boulay, B., Luckin, R., & del Soldato, T. (1999). The plausibility problem: human teaching tactics in the 'hands' of a machine. In S. P. Lajoie & M. Vivet (Eds.), *Artificial Intelligence in Education, Open Learning Environments: New Computational Technologies to Support Learning, Exploration, and Collaboration, proceedings of AIED-99* (pp. 225–232). Amsterdam: IOS Press.

Ferguson-Hessler, M. G. M., & de Jong, T. (1990). Studying physics texts: differences in study processes between good and poor performers. *Cognition and Instruction, 7,* 41–54.

Flavell, J. (1979). Metacognition and cognitive monitoring. A new area of cognitive development inquiry. *American Psychologist, 34,* 906–911.

Gagné, R. M., & Smith, E. C. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology, 63* (1), 12–18.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9* (6), 495–522.

Griffin, S., Case, R., & Siegler, R. S. (1994). Rightstart: providing the central conceptual prerequisities for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice* (pp. 25–49). Cambridge, MA: MIT Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of Inference, Learning, and Discovery.* Cambridge, MA: The MIT Press.

Judd, C. H. (1908). The relation of special training to general intelligence. *Educational Review, 36,* 28–42.

Kalchman, M., Moss, J., & Case, R. (2001). Psychological models for development of mathematical understanding: rational numbers and functions. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 1–38). Mahwah, NJ: Erlbaum.

Katona, G. (1940). *Organizing and Memorizing.* New York: Columbia University Press.

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: elements of expertise in geometry. *Cognitive Science, 14,* 511–550.

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™: Summary Research Results.* White paper. Available from Carnegie Learning Inc., 1200 Penn Avenue, Suite 150, Pittsburgh, PA 15222, E-mail: info@carnegielearning.com, Web: http://www.carnegielearning.com

Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: an updated analysis. *Computers in Human Behavior, 7,* 75–95.

Lepper, M. R., Woolverton, M., Humme, D. L., & Gurtner, J. L. (1993). Motivational techniques of expert human tutors: lessons for the design of computer-based tutors. In S. Lajoie & S. Derry (Eds.), *Computers as Cognitive Tools* (pp. 75–105). Hillsdale, NJ: Erlbaum.

Lovett, M. C. (1992). Learning by problem solving versus by examples: the benefits of generating and receiving information. In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society* (pp. 956–961). Hillsdale, NJ: Erlbaum.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: the effect of example format and generating self-explanations. *Cognition and Instruction, 16,* 173–199.

NCTM. (1989). Curriculum and Evaluation Standards for School Mathematics. National Council of Teachers of Mathematics. Reston, VA: The Council.

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer, memory-oriented training promotes memory for training. *Memory and Cognition, 19,* 543–557.

Paivio, A. (1971). *Imagery and Verbal Processes.* New York: Holt, Rinehart, and Winston.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension monitoring activities. *Cognition and Instruction, 1,* 117–175.

Renkl, A. Learning from worked-out examples: instructional explanations supplement self-explanations. *Learning and Instruction* (in press).

Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science, 21,* 1–29.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: the effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23,* 90–108.

Rose, C. P., & R. Freedman (Eds.). (2000). *Building Dialogue Systems for Tutorial Applications. Papers from the 2000 AAAI Fall Symposium.* Menlo Park, CA: AAAI Press.

Scardamalia, M., Bereiter, C., & Lamon, M. (1996). The CSILE Project: trying to bring the classroom into world 3. In K. McGilly (Ed.), *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice* (pp. 201–228). Cambridge, MA: The MIT Press.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive Science and Mathematics Education* (pp. 189–215). Hillsdale, NJ: Erlbaum.

Schofield, J. (1995). *Computers and Classroom Culture.* Cambridge University Press.

Simon, H. A. (1987). The information-processing explanation of Gestalt phenomena. In *Models of Thought, Volume II* (pp. 481–493). New Haven, CT: Yale University Press. (Reprinted from *Computers in Human Behavior, 2,* 1–15, 1987.)

Schwartz, J. L., Yerushalmy, M., & Wilson, B. (Eds.). (1993). *The Geometric Supposer: What is it a case of?* Hillsdale, NJ: Erlbaum.

Singley, K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *The Journal of The Learning Sciences, 2* (10), 1–59.

Wiske, M. S., & Houde, R. (1993). From recitation to construction: Teachers change with new technologies. In J. L. Schwartz, M. Yerushalmy, & B. Wilson (Eds.). *The Geometric Supposer: What is it a case of?* Hillsdale, NJ: Erlbaum.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Education Research, 13,* 21–39.

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge.* Los Altos, CA: Morgan Kaufmann.

Wertheimer, M. (1959). *Productive Thinking.* New York: Harper & Row.

White, B. Y., Shimoda, T. A., & Fredericksen, J. R. (1999). Enabling students to construct theories of collaborative inquiry and reflective learning: computer support for metacognitive development. *International Journal of Artificial Intelligence in Education, 10* (2), 151–182.