To appear in: *Proceedings of the AAAI 2000 Fall Symposium, Building Dialogue Systems for Tutorial Applications.* Menlo Park, CA: AAAI Press.

The Need for Tutorial Dialog to Support Self-Explanation

Vincent A.W.M.M. Aleven and Kenneth R. Koedinger

Vincent A.W.M.M. Aleven and Kenneth R. Koedinger
HCI Institute
School of Computer Science
Carnegie Mellon University
aleven@cs.cmu.edu, koedinger@cs.cmu.edu

Abstract

What are good ways of using natural language dialog in intelligent tutoring systems? A role with high potential pay-off is to support the meta-cognitive process of self-explanation. In previous experiments involving the PACT Geometry Tutor, we found that students learn with greater understanding, when they are required to explain their solutions steps "by reference", that is, by naming the rule that was used. However, the tutor may be even more effective if students explain their solution steps in their own words and if the tutor helps them, through dialog, to improve their explanations. An exploratory experiment with a tutor version that did not do any natural language processing, strongly suggested the need for natural language dialog. Without feedback from the tutor, students provided few free-form explanations in response to the tutor's prompts. Of the explanations that they did provide, only a small portion were correct and complete. During the experiment, we also identified a number of dialog strategies that we plan to implement in the tutor.

Introduction

Natural language processing technologies have matured enough that we can begin to apply them in educational software. But how can we use these technologies most effectively? What kind of natural language dialog will have the greatest pedagogical leverage? In this paper we present our explorations of these questions in the context of the PACT Geometry Tutor, a cognitive tutor for high school geometry. Like many other 2nd generation tutoring systems developed in the 80s and 90s, the PACT Geometry tutor helps students learn to *apply* knowledge, by providing opportunities for guided learning by doing, with context-sensitive feedback and hints. Empirical studies have confirmed that practice with such 2nd generation tutors is effective [Koedinger, *et al.*, 1997]. Nonetheless, there is room for improvement. For example, there is evidence that intelligent tutoring systems are not as effective as human one-on-one tutors [Bloom, 1984].

The mission of the CIRCLE Research Center, with which the authors are affiliated, is (in part) to develop 3rd-generation tutors, which help students to *construct* knowledge and tutor metacognitive processes that are known to improve learning. One type of metacognitive skill is self-explanation. A number of cognitive science studies show that students learn better to the extent that they explain the materials to be learned to themselves [Chi, *et al.*, 1989; Bielaczyc *et al.*, 1995]. Self-explanation serves as a check on one's understanding. When students are unable to construct an explanation, this means that they have detected a gap in their knowledge. If they can fill in the gap, new knowledge is constructed [VanLehn, *et al.*, 1992]. However, this process requires metacognitive skill: Students must be able to detect when an explanation is inadequate, and must be able to fill gaps that are detected in a meaningful way. Students differ in their ability to do so [Chi, *et al.*, 1989].

Prompting students to explain learning materials helps to elicit effective self-explanations and leads to greater learning [Chi, *et al.*, 1994; Renkl 1997]. Also, self-explanation skills can be taught through an instructional program that included one-on-one instruction with a human instructor [Bielaczyc, *et al.*, 1995]. While these studies suggest that instructional programs that emphasize self-explanation can be effective, they leave open the question how self-explanation can be supported in an actual classroom. Neither prompting nor a program of one-on-one instruction seem feasible in this regard. Further, it is not clear how effective these interventions would be in the classroom. Bielaczyc's instructional program works for students enrolled in one of the nation's elite universities, but does it follow that it works also for average high-school students? Further, prompting does not benefit all students as much as would be ideal [Renkl, 1997]. In order to get *all* students to produce effective self-explanations, students need more continuous guidance and feedback.

Therefore, like [Conati, 1999], we focus on self-explanation as a way of improving a 2nd generation tutor. In our previous research, we found that a tutor version that required students to explain their solution steps "by reference" leads to greater understanding [Aleven, *et al.*, 1999]. In this paper, we argue that in order to take the next step, it is important that students explain in their own words, rather than by providing references, or using a structured interface. Further, we present results from a study that we conducted to find out whether natural language understanding and dialog are really needed or whether it is sufficient that students explain in their own words, without receiving feedback from the tutor on their explanations. In this study, we tested a tutor that prompts students to give explanations, but does not analyze or check the correctness of these explanations ("unchecked NL explanation"). *A priori*, it seemed conceivable that such a tutor could be more effective than one that does not prompt students to provide explanations. After all, this tutor is in many ways similar to the unguided or prompted self-explanation conditions in the self-explanation studies mentioned above, which were shown to improve learning. Conversely, if a tutor that supports unchecked NL explanations would turn out not to be very effective, this would provide evidence that dialog will have to be a crucial ingredient of the 3rd generation of intelligent tutoring systems.

Adding support for explanation to a 2nd generation tutor

The PACT Geometry Tutor helps students learn skills of geometry problem-solving. Like all cognitive tutors, it supports guided learning by doing [Anderson, *et al.*, 1995]. The PACT Geometry Tutor is currently in use in four schools in the Pittsburgh area and in one school in Kentucky. The tutor is part of a full-year high-school geometry course, in which students spend 40-50% of the classroom time working problems on the computer. The tutor curriculum consists of six units: Area, Pythagorean Theorem, Angles, Similar Triangles, Quadrilaterals, and Circles. In most tutor problems, students are presented with a diagram, and are asked to find unknown quantities, such as angle or segment measures. The tutor has a detailed cognitive model, in the form of production rules, of the geometry problem-solving skills that are needed to solve these kinds of problems. Using its cognitive model, the tutor monitors students as they enter solution steps, gives hints on demand, and provides feedback. The tutor's cognitive model is also the basis student modeling. The tutor keeps track of how well the student masters each skill in the model and uses that information to support mastery learning.

We conducted a formative evaluation of an early version of the PACT Geometry Tutor, which required students to solve problems but not to explain their solution steps. We found significant learning gains, attributable to the combination of working on the tutor and classroom instruction [Aleven, *et al.,* 1998]. But we also found some evidence of shallow learning: students were not always able to explain their answers even if the answer itself was correct. This we attributed to the use of shallow heuristics, such as: if two angles look the same, they are the same. This motivated the next step.

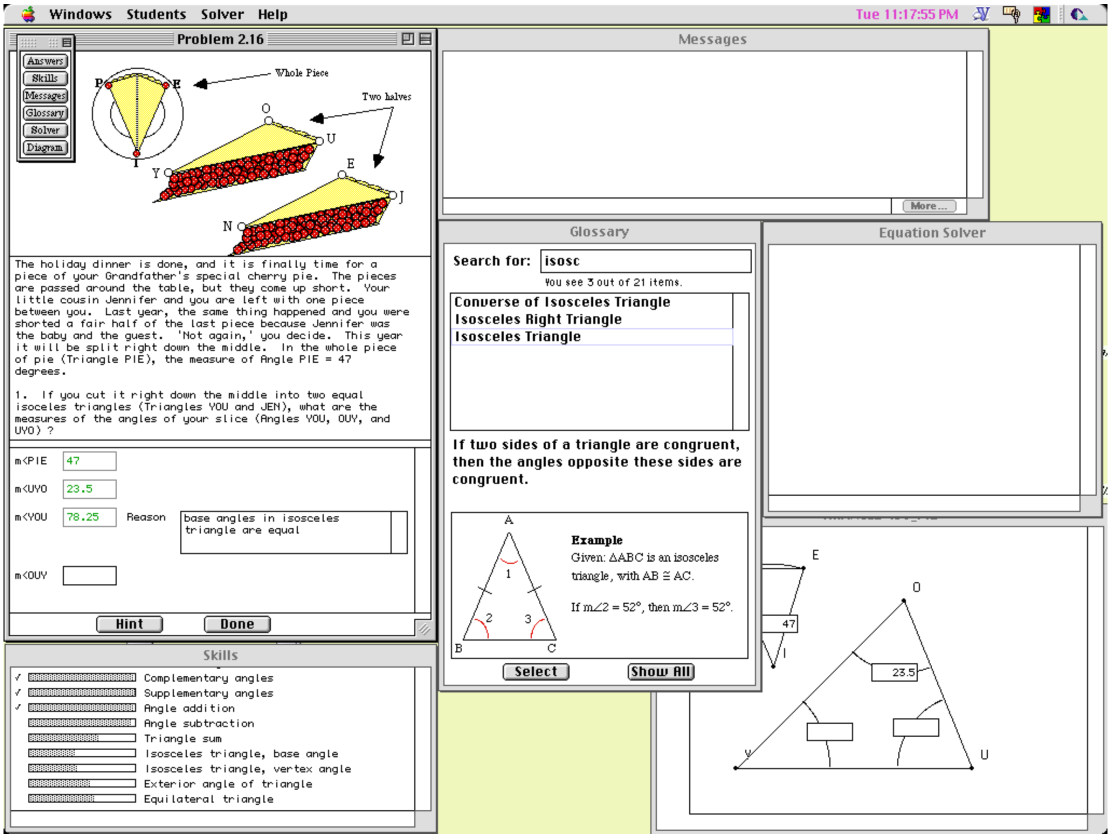
As part of our effort to move from the 2nd generation of tutors to the 3rd, we then constructed a version of the PACT Geometry Tutor which, without relying on any NL dialog capabilities, required students to provide correct explanations for solution steps. Students could enter explanations by providing a reference, that is, by naming the geometry rule or definition that justifies the step. The tutor presented all relevant rules in a separate Glossary window on the screen, with a description of each rule and an example. Students could select the reference from the Glossary, or could type it in. Two evaluation studies provided evidence that students learned with greater understanding when they explain answers by reference. Students who had explained their answers during training were better able to provide reasons for solutions steps, and were better able to deal with transfer problems [Aleven, *et al.*, 1999]. These are typical measures of understanding.

Why a 3rd generation tutor is likely to be better

While explanation by reference is effective, there is good reason to believe that a tutor would be even more effective if students provide full-blown explanations rather than references, and if they state explanations in their own words. A full-blown explanation would be, for example, a statement of a general domain rule or principle. Further, the explanation could indicate how that rule was applied to the problem at hand. Students may learn more when they give full-blown explanations, because this forces them to rehearse the domain rules or principles more fully than when they provide a reference only. Thus, it may lead to better learning. One way for students to provide full-blown explanations, is through a structured editor, or a template-based interface, akin to that of the SE-COACH, an intelligent tutoring system that helps students explain examples [Conati, 1999]. An advantage of a structured interface is that it provides a certain level of scaffolding: It allows only a certain class of explanations to be entered and thereby communicates constraints on what is a good explanation.

Nonetheless, it may be even better if students can state their explanations in their own words. First, natural language is natural. There is no need to learn a new interface. More importantly, when students explain in their own words this makes it easier to build on their partial knowledge. Students can give an explanation based on what they know (or do not know) and go from there. There is no need to translate what they know into the language of a structured editor, for example. In that translation step, partial knowledge may not help.

Further, there is a recall v. recognition issue. When students explain in their own words, this forces them to recall the relevant knowledge from memory. On the other hand, when they use a structured editor or menus, they rely less on recall from memory and more on recognition. Usually the interface is structured in such a way that one can piece together an explanation by recognizing, in a menu or list, the correct choices for various components of



The holiday dinner is done, and it is finally time for a piece of your Grandfather's special cherry pie. The pieces are passed around the table, but they come up short. Your little cousin Jennifer and you are left with one piece between you. Last year, the same thing happened and you were shorted a fair half of the last piece because Jennifer was the baby and the guest. 'Not again,' you decide. This year it will be split right down the middle. In the whole piece of pie (Triangle PIE), the measure of Angle PIE = 47 degrees.

1. If you cut it right down the middle into two equal isoceles triangles (Triangles YOU and JEN), what are the measures of the angles of your slice (Angles YOU, OUY, and UYO) ?

m<PIE 47

m<UYO 23.5

m<YOU 78.25 Reason base angles in isosceles triangle are equal

m<OUY

Search for: isosc

You see 3 out of 21 items.

Converse of Isosceles Triangle
Isosceles Right Triangle
Isosceles Triangle

If two sides of a triangle are congruent, then the angles opposite these sides are congruent.

Example
Given: ΔABC is an isosceles triangle, with AB ≅ AC.

If m∠2 = 52°, then m∠3 = 52°.

Figure 1: The unchecked NL version of the PACT Geometry Tutor.

the explanation. But recall is superior to recognition in terms of its learning benefits.

Finally, when students explain in their own words, there is less of a chance that students have problems with jargon or unfamiliar terminology. Or at least, if they do not know a certain term, they can describe the concept in their own words.

Exploratory study - paving the way for the 3rd generation

Before developing a dialog system, we wanted to have a better understanding of whether dialog would really be needed. Therefore, we conducted a study to find out how effective a system is that supports unchecked NL explanation. In other words, we wanted to see what happens when students work with a tutor that provides for guided learning by doing (with feedback), prompts students to type explanations of their solution steps, but does not give feedback on explanations. It seemed plausible that even without feedback, NL explanation would have a positive effect on students' learning. At first blush, the unchecked NL condition is not that much different from the (prompted or unprompted) self-explanation conditions in the Chi *et al.* studies. If these conditions could lead to learning in the absence of feedback, then why not unchecked NL explanation in a computer tutor? If, on the other hand, unchecked NL explanation turned out not to be very effective, than this would clearly suggest the need for natural language understanding and dialog.

A second goal of the study was to find out more about the task that a dialog system would face, if it were to support self-explanation in the geometry domain. We were very interested to find out more about the kinds of explanations that students type, and to find out how good these explanations are. Obviously, that kind of information is important for developing an NLP component for a tutoring system. Finally, we wanted to know if students would type explanations at all, if this is not strictly mandatory. Clearly, there was a risk that as soon as word got out that the tutor does not check the explanations, students would stop typing them altogether.

The study was carried out in a high school in the Pittsburgh area, where the PACT Geometry Tutor is used as part of the regular geometry instruction. The study involved 36 students in two periods taught by one teacher. All students completed the unit of the tutor curriculum that deals with angles. The students also received classroom instruction on the topics of that unit. At the start of each tutor problem, the tutor selected the steps that it would prompt the student to explain. The selection mechanism, which involved a random element, was designed to make sure that students were asked to explain only a small number of steps per problem. The tutor prompted students to explain steps simply by placing a box for typing the explanation next to the corresponding answer box (see Figure 1, window on the top left).

Results

We analyzed students' explanations, extracted from the logs of the students' sessions with the tutor. These logs, collected automatically, record in considerable detail student actions as well as the tutor's responses. We classified the students' explanations with respect to eleven analysis categories, illustrated in Table 1. First, we distinguished two types of *correct explanations:*

Complete Explanations The student gave a correct and complete explanation. This includes statements of geometry rules, even if they do not say how the rule was applied to the problem at hand, and even if the language is inexact or disjointed. Explanations that state how a general rule was applied, without stating the rule itself, also count as Complete Explanations. We were very lenient in our assessment of students' explanations, so as not to underestimate students' ability to self-explain.

Reference This category includes references to a geometry theorem or definition, presented without any elaboration. The reference has to be to a rule that can actually be used for the given step.

We also identified a number of categories of *incorrect or partially correct explanations.*

Procedural Replay The student did not provide a justification for the solution step in terms of domain rules and principles, but merely restated the answer itself, or the arithmetic operations involved in finding the answer.

Incomplete Explanation The student attempted to state a general rule of geometry, but left out part of the rule, usually some of the conditions.

Equal Angles The student stated that an unknown angle measure was equal to the measure of another angle in the diagram, without further justification. This kind of explanation can be seen both as a Procedural Replay and an Incomplete Explanation, which is why we created a separate category.

Incorrect Reference The student gave a reference to a geometry rule that was not applicable to the given step.

Table 1: Examples of students' explanations with the unchecked NL version of the PACT Geometry Tutor

Correct Explanations

Complete Explanation
A complementary is when two equal 90 degrees
the maeasure of a straight line is 180 so you just take the number left and subtract it
vertical angles so that makes them the same
because it is an equalateral triangle, so you take 180 and divid it by 3
Reference
because it a isocles triangle
triangley stuff [wordplay on triangle sum]

Partially-Correct or Incorrect Explanations

Procedural Replay
180 minus 160.6 is equal to 19.4
take 71 given then multiply that by 2 then subtract from 189 71+71=142 then subtract that from 180
I took the measure of angle DRO and ORPand added them to togrther
Incomplete Explanation
all angels equal 180
angle 1 and angle 6 equal 90 degrees
they are parallel to each other
Equal Angles
angle ARP and angle IRL are identical
all have the same measure
Almost Correct Explanation
because of the parrel lines in the (-) of the 180 [supplementary interior angles]
the angles all branch off of the same center point [vertical angles]
the area in between the two triangles is those two numbers added together [angle addition]

Non-Attempts

Non-Explanation
i dunno
Because.......... i said so..
I USED HELP..…
Off-Task Comment
I am god
math told me, he's a little lepracon that lives in my ear, him and all his friends.
cause i am ready for lunch

Given The student tried to justify a step by saying that the information was given in the problem statement. Such explanations are incorrect because the tutor did not ask students to explain given steps.

Almost Correct Explanations This category includes explanations that did not fit any other categories.

Finally, we identified three categories of *non-attempts*

Blank The student did not provide any explanation.

Non-Explanation This category includes inputs with little explanatory value such as "I just know".

Off-Task Comment This category includes amusing comments not related to the study of geometry, that is, Non-Explanations but with higher entertainment value.

The protocol data indicate that the tutor prompted students to type an explanation for 22% of the steps (1234 out of 5615). This amounts to 34±17 explanations per student. Students attempted an explanation on 36% of these steps (439 steps)—see Table 2. The rest of the time, the explanation boxes were left blank or filled with comments of varying degrees of interestingness and X-ratedness (the categories of Non-Attempts listed in Table 1). Students entered a correct explanation on 9% of the steps, or equivalently, on 26% of the steps for which they actually tried to produce an explanation. More than half of the correct explanations were references, not full-blown explanations. The vast majority of students (23 out of 36) did not explain a single geometry rule correctly (see Figure 2). The highest number of different geometry rules explained correctly by any given student was five (see Figure 2). The number of different rules that the tutor asked each student to explain was 11±1.8. Thus, no student explained more than half the assigned rules correctly and the average was far below that. The categories of partially correct or incorrect explanations that occurred most frequently were Procedural Replay, Equal Angles, and Incomplete Explanation.

Table 2: frequency of the 11 explanation categories (percentage of the total number of solution steps for which an explanation was required)

Correct Explanations	**9%**	
Reference		6%
Complete Explanation		4%
Partially-Correct or Incorrect Explanations	**26%**	
Procedural Replay		10%
Equal Angles		5%
Incomplete Explanation		4%
Incorrect Reference		4%
Almost Correct Explanation		2%
Given		1%
Non-Attempts	**64%**	
Blank		41%
Non-Explanation		19%
Off-Task Comment		4%

Discussion

The results of the exploratory study indicate that explaining geometry rules is difficult for students. the number of correct explanations was low, even when using a very lenient criterion for correctness. Only 13 out of 36 students explained at least one rule correctly and completely (see Figure 2). No students correctly explained more than half the number of different rules that they were asked to explain. Students provided a correct explanation for only 26% of the steps where they actually made an attempt at explaining, or 10% if we discount explanations that are mere references. While references may be adequate in some situations (e.g., for simpler skills or for skills that the student has mastered), it is better if students give full-blown explanations, as we have argued.

The results indicate further that students often did not comply with the tutor's prompts for explanations. Students attempted to explain their answer not much more than one-third of the time they were asked to do so. Thus, working with a computer tutor, students do not seem to be inclined to respond to prompts for explanations very much, if the tutor does not analyze the content of the explanations.

The informal observations were that students often did not know what kind of explanations were expected. In all fairness, this was not explained to them beforehand. The experimenter tried to help, but more may be needed. It seems likely that students would provide more and better explanations if the students received some form of explanation training in the classroom (i.e., without the tutor) or were shown examples of the types of explanations that we would like to elicit. It might also help if students were told that the teacher would check the explanations. On the other hand, during the sessions with the tutor, the first author often tried to get students to improve their explanations. This was usually successful, but not always easy easy. If these experimenter/student dialogs are any indication, it may take continued guidance to get students to provide good explanations.

Further, students often did not appear to see the point of giving reasons for their answers. It may not be clear to most of them that this is a good study habit—a way of checking their own understanding. Many students seemed to have the impression that after giving the right answer, as indicated by the tutor's feedback, there was nothing left to explain. This lack of interest in explaining and self-

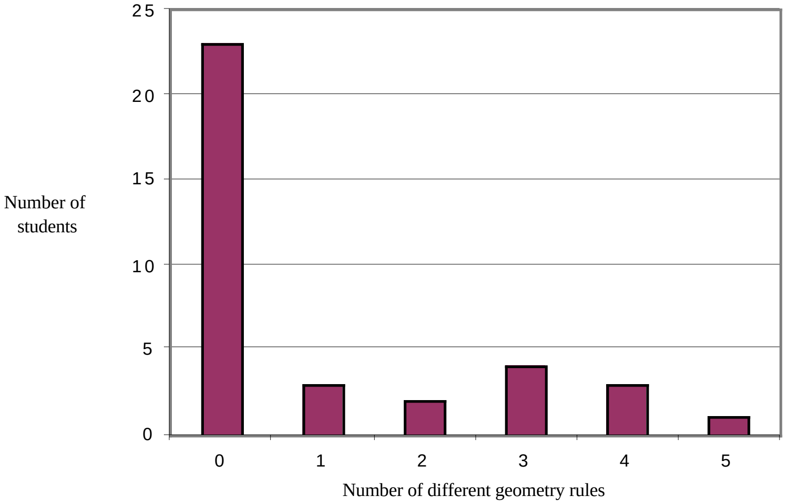


Figure 2: Number of students who explained a given number of geometry rules correctly (i.e., provided at least one Complete Explanation). The number of different rules to be explained was different for each student. The average was 11 different rules, the maximum was 13.

monitoring is a significant finding in itself. It suggests that students tend to minimize the amount of work needed per tutor problem, rather than trying to maximize the learning per problem. It suggests also that they may not have the metacognitive skills required for self-explanation.

In sum, very few students produced effective self-explanations with any degree of consistency. While there is some reason to think that unchecked NL explanation could work better, if it was better integrated with classroom instruction, it seems that more instructional guidance is needed. In particular, it would be very useful if the tutor could provide somewhat detailed feedback on explanations, beyond yes or no, and could conduct a dialog to help students improve their explanations.

Designing the 3rd generation: How many ways can you say “triangle sum”?

The explanations collected during the study provide insight into the kinds of inputs that a natural language understanding system for the geometry domain will have to handle. For example, the triangle sum rule, one of the rules that the students were asked to explain, was stated in the tutor's Glossary as follows:

> The sum of the measures of the three interior angles of a triangle is equal to 180°.

When students express geometry rules in their own words, we see a striking variety of language use, including many interesting variations and shorthand, as compared to the "official versions" of the rules. The complete set of explanations of the triangle sum rule that were deemed correct and complete is as follows:

- A triangle is comprised of 180 degrees
- because a triangle's sum adds up to 180 so i just subtracted the sum of the two given triangles
- triangle equals 180
- 180 degrees in a triangle
- Interior angles in a triangle equal 180
- IN THE TRIANGLE THE THREE ANGLES EQUAL T0 180 DEGREES.
- triangle add up to 180
- because all the angles in a triangle equal 180, and you add the two angles that you have and subtract
- because all the angles of a triangle equal 180

One challenge for a tutorial NLU system is to achieve sufficient coverage, in the face of this diversity. What should count as a correct statement of each geometry rule is subject to debate. In the current study, we were very lenient. We considered as correct all explanations that seemed to express the right idea. Some explanations that were deemed correct would probably not pass muster with a human teacher. A tutoring system might also use a somewhat stricter criterion for correctness, or better yet, different criteria depending on the situation and student. But regardless of the criterion used, the NLU system had to deal with a diverse set of inputs for any given rule to be explained.

Further, the students' explanations are characterized by loose use of language, involving among other things

- unusual choice of vocabulary, such as the use of the word "comprises" to mean the sum (first explanation of the triangle sum rule listed above);
- the occasional wrong choice of technical terms, for example saying "triangle" when "angle" is meant (see the second explanation of the triangle sum rule), or using the term "area" when angle measures are meant (see the third example in category Almost Correct Explanation in Table 1), or using the term "sides" when talking about angles (as we observed in the protocol data);
- abundant use of metonymy and other abbreviation phenomena, such as saying that the "triangle adds up to 180" when it is really the measures of the angles of the triangle that add up to 180.

The NLU component in educational systems will have to be robust in the face of such loose or unusual language use. It is important that the NLU component can come up with enough of an analysis of a sentence like "A triangle is comprised of 180 degrees" that the tutor does not have to reply: "Sorry, your explanation does not make sense to me, can you give it a good semantic make-over?" A much more productive reply is: "you have the right idea, but what exactly is 180 degrees?" Another possible reply might be: "What do you mean, a triangle comprises 180 degrees? A triangle comprises three sides and three angles." It is not hard to imagine that such replies would be followed by productive dialog that helps the student to state the triangle sum rule more precisely.

A companion paper describes the progress that has been made on developing a NLU component that is capable of analyzing students' explanations [Popescu and Koedinger, 2000].

Designing the 3rd generation: dialog strategies to help students improve their explanations

The exploratory experiment suggests that in order to support effective self-explanation, a tutoring system needs to be able to analyze students' explanations and engage students in natural language dialog. Developing such a dialog system is a significant technical challenge. As a first step in designing such a system, we discuss what kind of dialog strategies would be useful to help students improve their explanations. We also note what constraints are implied for the NLU component. The dialog strategies are loosely based on actual dialogs that the first author had with students, as they were working with the unchecked NL explanation version of the tutor.

As a general strategy, the tutor should try to get the student to state a general geometry rule, if the student did not try to do so. This strategy is appropriate when the student gave a correct reference (Reference), or a procedural replay (Procedural Replay), or stated that the angle of interest is equal to another angle (Equal Angles). For example, when the student states that angles are equal (Equal Angles), the tutor could ask to state the general rule:

S: Angle YGR is equal to angle XGH.
T: You are right about that, but how do you know that they are equal? Can you state a general rule about angles that you used?

When the student gives a Procedural Replay, the tutor could ask the student why that was correct, for example by using a *reductio* strategy, similar to entrapment strategies described in Collins' and Stevens' theory of Socratic tutoring [Collins and Stevens, 1982]:

S: I subtracted 79 from 180.
T: Whenever you have 79, do you always subtract it from 180?
S: No.
T: OK. So why did you use 180 in this step? Can you state a general geometry rule that tells you that 180 is the number to use here?

If the student tries to state a general rule, but is not successful, then the tutor could try a similar *reductio* strategy, in order to help the student arrive at a more complete explanation. For example, when the student explains the linear pair theorem (i.e., the sum of the measures of two adjacent angles that together form a line is 180 degrees) the following dialog might ensue:

S: The two angles are 180 degrees.
T: You are right, but how do you know this? Whenever you have two angles, is the sum of their measures always 180 degrees?
S: No.
T: Good. What is it about these two angles that tells you that the sum is 180 degrees? Can you state a general geometry rule?
S: Well, they form a line.
T: OK, good. Can you now state the general rule?
S: A line is 180 degrees.

At this point in the dialog, it is clear that the student has the right idea, but she has not yet stated a rule with sufficient precision. The tutor's goal is to get her to say "The measures of adjacent angles that together form a line add up to 180 degrees" or something similar, without dragging on the dialog for too long. An effective move might be:

T: Yes. So now let us work on stating a general rule. Please fill in the blanks: When two angles ________, you know that __________ is 180 degrees.
S: When two angles form a line, they are 180 degrees.
T: Excellent. Remember that rule because you will need it later.

The tutor's task is more difficult when students get a solution step right but cannot justify their answer. If the student focused on the wrong rule (perhaps gave an Incorrect Reference), a good dialog strategy might be to ask the student to state that rule, and then to ask her to explain how that rule applies to the current problem. That explanation is of course bound to fail, since the rule does not apply, but in the process the student is likely to see what rule is the right one to use. In situations where the student does not know which rule applies, a good dialog strategy identifies plausible candidates, and for each one, discusses if it applies. The plausible candidates can be found for example by searching a resource like the Glossary. However, these dialog strategies are too sophisticated for a first version of the tutor. At least for the time being, these situations are probably better handled through other facilities, such as on-demand help. Currently, the on-demand hints of the PACT Geometry Tutor are designed to communicate this kind of strategy.

We plan to use the Atlas planner to implement these strategies [Freedman, 1999]. The main requirement for the NLU component of the system is that it can recognize, not only when an explanation is a correct statement of a rule, but can also recognize various subcategories of partially correct or incorrect answers. For example, in order to implement the *reductio* strategy discussed above, the NLU component must be able to classify an explanation as a Procedural Replay. Or it must be able to recognize specific subcategories of Incomplete Explanations, such as statements that “the three angles are 180 degrees.” Moreover, it is probably good if the tutor ignores all Non-Explanations and Off-Task Comments. This means it must be able to recognize all other types of inputs, that is, all attempts at explanations. For further information, see the paper by Popescu and Koedinger [2000].

Conclusion

Our goal is to develop a 3rd-generation intelligent tutoring system, with improved effectiveness, as compared to current systems. In previous studies, we found that students learn with greater understanding if they explain their solution steps by reference, that is, by naming the problem-solving principle that was applied. However, students may learn even better if they explain their solution steps in their own words and state more complete explanations. This has the advantage that students have to attend to the relevant knowledge more fully, can build on existing partial knowledge more easily, and are forced to rely on recall of correct knowledge, not just recognition.

To find out whether natural language understanding is crucial, we evaluated a tutoring system that prompted students to type explanations in their own words, but gave no feedback on the explanations. Since this unchecked NL explanation is not unlike self explanation conditions found to be effective in other studies, there was a distinct possibility that this by itself would be effective. The results of the study suggest however that unchecked NL explanation leaves considerable room for improvement. When the tutor does not check explanations for correctness, students provide few explanations. Moreover, of the explanations that they provide, only about 26% are correct, and only about 10% are correct and not mere references.

The tutor may be far more effective if it could analyze students' explanations and provide feedback, and engage students in NL dialog in order to help them improve their explanations. The data from the exploratory study indicate what kinds of incorrect and partially correct explanations the system will have to deal with, and what the relative frequency is of those categories. We have identified a number of dialog strategies that were effective in the hands of a human tutor. Further, we have identified some constraints that are placed on the system's natural language understanding component. We plan to implement these strategies in our tutoring system and show that natural language dialog to support self-explanation has high pay-off.

References

Aleven, V., K. R. Koedinger, and K. Cross, 1999. Tutoring Answer Explanation Fosters Learning with Understanding. In *Artificial Intelligence in Education, Open Learning Environments: New Computational Technologies to Support Learning, Exploration, and Collaboration, proceedings of AIED-99*, edited by S. P. Lajoie and M. Vivet, 199-206. Amsterdam: IOS Press.

Aleven, V., K. R. Koedinger, H. C. Sinclair, and J. Snyder. Combatting shallow learning in a tutor for geometry problem solving. In *Intelligent Tutoring Systems,Fourth International Conference, ITS '98,* edited by B. P. Goettl, H. M. Halff, C. L. Redfield, and V. J. Shute, 364-373. Lecture Notes in Computer Science 1452. Berlin: Springer Verlag, 1998.

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

Bloom, B. S., 1984. The 2 sigma problem: the search for methods of group instruction as effective as one on one tutoring. *Eudcational Researcher,* 13, 4-16..

Bielaczyc, K., P. L. Pirolli, and A. L. Brown, 1995. Training in Self-Explanation and Self-Regulation Strategies: Investigating the Effects of Knowledge Acquisition Activities on Problem Solving. *Cognition and Instruction,* 13 (2), 221-252.

Chi, M. T. H., M. Bassok, M. W. Lewis, P. Reimann, and R. Glaser, 1989. Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science,* 13, 145-182.

Chi, M. T. H., N. de Leeuw, M. Chiu, and C. Lavancher, 1994. Eliciting Self-Explanations Improves Understanding. *Cognitive Science*, 18, 439-477.

Collins, A, and A. L. Stevens, 1982. Goals and Strategies for Inquiry Teachers. In *Advances in Instructional Psychology, Vol II*, edited by R. Glaser. Hillsdale NJ: Erlbaum.

Conati, C., 1999. An Intelligent Computer Tutor to Guide Self-Explanation While Learning from Examples. Ph.D. Thesis, Intelligent Systems Program, University of Pittsburgh.

Freedman, R., 1999. Atlas: A Plan Manager for Mixed-Initiative, Multimodal Dialog. *AAAI-99 Workshop on Mixed-Initiative Intelligence*, Orlando, FL.

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.

Popescu, O., and K. R. Koedinger, 2000. Towards Understanding Geometry Explanations In *Proceedings of the AAAI 2000 Fall Symposium, Building Dialog Systems for Tutorial Applications.* Menlo Park, CA: AAAI Press.

Renkl, A., 1997. Learning from Worked-Out Examples: A Study on Individual Differences. *Cognitive Science* 21 (1), 1-29,

VanLehn, K., R. M. Jones, and M. T. Chi, 1992. A Model of the Self-Explanation Effect. *The Journal of the Learning Sciences,*, 2 (1), 1-59.