WHAT HAPPENS IF YOU INTRODUCE AN INTELLIGENT TUTORING SYSTEM IN THE CLASSROOM: A CASE STUDY OF THE GEOMETRY TUTOR

Yasmin Kafai
Department of Electrical Engineering and Computer Science
Advanced Technologies Laboratory
University of Michigan

In the last decade several Intelligent Tutoring Systems have been developed. The goal of Intelligent Tutoring Systems (ITS) is to help people acquire expertise through individualized learning. These systems have reached a level of functioning which allows evaluation of their effectiveness. This report describes the introduction of an intelligent tutoring system into a classroom setting. We used the Geometry Tutor developed by Anderson et al. (1985) at Carnegie Mellon University in three classes to help the students do congruence proofs. We considered two major aspects in our evaluation: the transportability of the ITS into the classroom and the tutor's instructional effectiveness. The results of this study indicate that problems may arise in integrating an ITS into the curriculum. These problems have implications on the instructional effectiveness of intelligent tutoring systems. However, most students volunteered when working with the Geometry Tutor to do as many proofs as possible. Taking advantage of this aspect may indicate possible positive and empowering uses of the Geometry Tutor.

Introduction

Individualized learning has been shown to be more effective than methods of group instruction (Bloom 1984). This suggests that even suboptimal individualized pedagogical strategies may prove a more powerful educational advance when implemented via Intelligent Computer Assisted Instruction (ICAI). Various domains have been covered: Arithmetic skills (WEST Burton & Brown 1982, DEBUGGY Burton 1982), Geometry (Geometry Tutor Anderson, Boyle & Yost 1985), Programming (BRIDGE Bonar & Weil 1985, PROUST Johnson & Soloway 1986, LISP-Tutor Anderson & Reiser 1985), Electronics (QUEST White & Frederiksen 1985, SOPHIE Brown, Burton & DeKleer 1982), and Engineering (STEAMER Williams, Hollan & Stevens 1981). The approach taken is to create intelligent tutors or coaches to replicate many of the advantages of a private human tutor. Intelligent computer based tutors have the ability to work interactively, to follow what a student is trying to do, to diagnose difficulties the student is having, and to provide instruction relevant to that difficulty.

So far, most projects have concentrated on theoretical and technical issues of ICAI and have dealt more with the internal evaluation of these systems (Littman & Soloway 1987). Internal evaluation focusses on the relationship between the architecture of an ITS and its behavior. It tries to determine whether the ITS performs in the way the designers intended it to do or what knowledge the ITS incorporates. Little attention has been paid to the other important aspect of the evaluation process: the external evaluation. External evaluation seeks to determine the impact of the ITS on the students learning. Some of these systems have been exposed for a short time to 'real students' or in classroom settings (Table 1) with positive effects such as time savings in learning and better test results. But, only a few studies undertook such an external evaluation of their ITS (Baker et al. 1985, Sack et al. 1988).

One of our concerns in this paper is not only to undertake the external evaluation of one such an ITS, but also to broaden explicitly the definition of external evaluation. Besides the instructional effectiveness we additionally consider aspects such as the transportability and adaptability of an ITS in the classroom setting. Implicitly these issues have been part of the discussion, but treated under the seperate cover of pedagogical issues. We will demonstrate with our approach the need for an integrated evaluation.

Motivation and Goals

The system we evaluated was the Geometry Tutor (GT) designed by Anderson et al. (1985) at Carnegie Mellon University. Our choice was guided by practical motivations: there is a partial version of the GT available for the Macintosh and we had a classroom set up with nine Macintoshs SE's available. In this study it is not our concern whether the GT functions adequately on a theoretical level or if the GT is better than traditional methods of teaching. Our focus is to investigate whether effective learning with the tutor is possible and whether the GT is transportable to the classroom. We address the following issues:

Instructional effectiveness: One of the critical tests of an ITS is "Can it teach students what it is suppose to teach them?" An ITS stands or falls upon the ability to perform well in face of the pragmatic challenges of the classroom. We investigated the performance of the GT in instructing students and examined the students ability to use it effectively. Related to this aspect is that the GT uses a different proof method from textbooks to do proofs, i.e. the proof path (PP) method. The traditional method used in textbooks is called the statement-reason (SR) method. Were the students able to learn this new proof method when interacting with the GT? What were the specific problems of students using the proof path method? Additionally we integrated the students in the evaluation process. So we were able to analyze the students performance with the GT, and also to ask them to evaluate specific features of the GT.

Transportability of the system into the classroom setting: The issue of transportability includes such aspects as the teachers' lack of familiarity with the material and the machinery, logistical and practical impediments to computer usage, and the lack of guidelines for the integration of the systems. These aspects constitute actual problems in the classroom reality as Schofield and Verban (1987) indicated. Since the current technology is leaving its familiar surroundings (Dede 1986), unforeseen problems may arise. To explain this situation we might think of the different 'lives' an ITS has. The first life of the ITS deals essentially with the design of the system and making it run. This phase takes place in the laboratory. The second phase of the ITS concentrates on the instructional effectiveness. Here the system moves under the control of the research team into the classroom and students/users work with the system. The third and so far neglected life of the ITS takes place the moment when the system leaves the familiar surrounding and is available for other schools which are out of immediate reach of the research team. Here the teacher takes the role previously played by the research team and has to handle the machines, provide introduction and fit the machines into his/her curriculum.

The questions we tried to answer: how can these systems be introduced into the traditional classroom setting, and what are the problems which one may encounter from teachers and students? What kind of introduction method need to be supplied? How does the system fit conceptionally into the curriculum of geometry classes? What are the instructional differences (textbook vs tutor)? We provided an introduction method, the workshop, which seemed to be appropriate for the present version of the GT.

We start with a description of the Geometry Tutor. The next step is to describe the study we undertook at a public high school with three geometry classes - two normal and one advanced. We then present an overview of how we introduced the Geometry Tutor and the results of that introduction. Finally we examine and discuss the benefits of the GT.

Description of the Geometry Tutor

The Geometry Tutor (GT) was designed by Anderson, Boyle, & Reiser (1985) to assist high school students in constructing proofs in Euclidian plane geometry. A number of principles directed the construction of the Geometry Tutor. We will discuss two of these principles because of their educational relevance here: the use of the proof path, and the instruction in context/immediacy of feedback.

Use of the Proof Path: According to the authors, "*a major effort in the design of the interface has been to communicate to the student the logical structure of a proof and the structure of the problem solving process by which a proof is generated* " (Anderson et al. 1985, p.4). The result is a proof path presented on the screen (Fig.1,2,3,4) which includes the rules/definitions applied to the premises during the inferences in generating the proof. The student grows the path by a combination of pointing to statements on the screen and typing in information. The advantages of this method are that it illustrates the structure of a complete proof and the critical features of the proof.

Instruction in Context and Immediacy of Feedback: Students appear to learn information better if they are confronted with information in the problem solving context (see Fig.1) rather than in instruction apart from the problem solving context. The GT provides immediate feedback on the students problem solving efforts. Whenever the students make a logically incorrect step of inference the system responds by identifying the logical error.

Different help facilities are available for the student:
- *Student seeking information*
Windows can be opened to see the set of rules and definitions covered so far and a more detailed explanation of these can be displayed on the screen.
- *Student seeking help*
In the command menu the **Explain** function allows the student to ask for additional help to continue the proof. Even if the chosen premises are not the most highly rated rule in the state of the proof by the tutor, the system will choose an adequate set which allows the student to continue. The student is offered at every trial a graduated hint (Fig.1). A graduated hint means to give the student help first in a very general way and to become more specific with every additional step.
- *Student getting help/information*
After three unsuccessful trials the tutor will do the step for the student.

A PC version of the GT is available on a Macintosh SE. It offers the student a set of 17 different congruence proof problems of varying difficulty. The GT provides help to the student through constant monitoring and structuring of the students problem solving attempts in the process of generating a proof.

In the following figures we give an example of how a student does a congruence proof in the Geometry Tutor. The student chooses one of the problems and receives on the screen the initial problem state (Fig.2).

The statement to be proven is displayed at the top of the screen, the givens are displayed at the bottom and a diagram in the upper left-hand corner. In the present version the student can only reason forward from the givens, applying rules and conclusions. Proof problems depending on their degree of complexity can require several proof steps to trace a set of inferences to the top (Fig. 3 represents an intermediate problem state). The final proof shows the complete path for the solution (Fig. 4) .

Introduction of the Geometry Tutor in the Classroom

The School and the Students

Three geometry classes (in total 27 students) of a public high school participated in this study. Two classes had average geometry students and one was a honors class in geometry. All classes had covered nearly the same material in geometry, essentially plane euclidian proofs of congruency, although the honors class was much farther through the curriculum. The geometry teacher, the same for of all classes, agreed to allow his classes to participate for one week at the Geometry Tutor Workshop during their normal class meeting.

The introduction to the Geometry Tutor did not take place in the classroom but in a workshop. One of the computer rooms in the school were set up with 9 Macintosh SE's. All of the computers were connected via a network system to a printer.

Description of the Geometry Tutor Workshop

The workshop was designed to make students familiar with the Macintosh and Geometry Tutor features, to enhance what students learn from normal geometry instruction and to teach new geometry rules not covered in the average geometry classes. The workshop was taught instead of the normal instruction and took place on five consecutive one period lessons in one week. The Geometry Tutor Workshop was divided into three different phases:

- the *Introduction Phase* when the students were taught how to use the Macintosh and the GT.
- the *Learning Phase* when every student worked at his/her own pace on a number of different proof problems and did homework either in the familiar statement-reason or in the proof-path method.
- The *Testing Phase* included a test for students, a questionnaire about computer attitudes, and an evaluation of the GT.

Some of the instruction and testing materials are presented in more detail along with the results. Further information can be found in Kafai (1988).

Results

The results are presented as follows: first we give an overview of the effects and impressions of the GT workshop and students. Then we evaluate the GT performance using two criteria: the instructional effectiveness and the transportability of the GT into the classroom.

Overall Performance

Summarizing the impressions, we can say that the use of the GT was a success in many regards for all three classes. In total 27 students participated in this study and all of them recommended the GT for other students. All students completed the minimal problem set and most of them did all of the additional 7 problems[1]. The students showed a surprisingly positive attitude towards computers regardless of their sex, class or geometry background. The average score on the computer attitude questionnaire for all students was 89 points (s=6.9) on a scale of 100 which indicates a fairly positive attitude. Apparently, computers are more familiar for students today. This is especially true in this school, since most students take computer related classes. We might further conclude that whatever the effects of the Geometry Tutor are, they will at least not be influenced by a negative attitude towards computers. Furthermore, the difference between the normal and the advanced geometry classes proved not to be significant. For this reason we decided to aggregate both data sets and all forthcoming analyses will refer to this one data set.

Instructional Effectiveness

Because the evaluation of the instructional effectiveness of the GT is particularly difficult, we approach this task from different sides. On one hand we evaluate the results of the actual performance of the student using the GT, on the other hand we use the students performances when applying the proof path method to do congruence proofs.

Homework

During the GT workshop we gave the students three assignments of two problems each to do at home. For every problem we told them what proof method to apply (see figure 5). One major characteristic of these homeworks was that all of the proofs had been done before by the student in class on the GT, so that the students were already familiar with the proof problem and structure. At the end of every proof we asked the students if they had problem doing the proof and which method they thought were easier for them to do the proof.

Looking at the percentages of correct solutions, we can see that the performance of the students decreases remarkably when using the proof path method. This low performance is related to the particular character of the tasks which will be described in more detail in section 'Transportability of the system". If the proof structure in the GT is not similar to the that of the statement-reason method, the students have more difficulties (86% and 62% in Homework 2 and 3). If the proof path is similar to the way the proof is done in the familiar statement-reason format, the difficulties decrease (29% in Homework 2). That the percentage is still relatively high might indicate the lack of familiarity with the proof method. No matter what method of proof the students were been assigned, 2/3 of them thought that the actual proof was easier to do in the statement-reason method.

Reconstruction of the Proof Process

A direct comparison between both proof methods: statement-reason, and proof path, can be drawn from the reconstruction task. Every student had 4 minutes to complete a simple geometry proof. The students in two classes (one normal geometry class and one advanced geometry class) were randomly assigned to two groups. One had to reconstruct a statement-reason solution, the other had to reconstruct the right order of the proof path

1 Remark from the Geometry teacher: "*It never happened to me that the students were asking for additional problems*."

(see Figure 6). In order to rearrange the proof path correctly the student had to draw arrows from the givens to the rule to apply and then to the conclusion.

As the results indicate the performance in the SR groups is significantly higher than that in the proof path method (Chi-Square = 20.16, df= 1, $p < 0.01$). We can conclude from that result that the proof task itself was not a problem, but the assigned method which refers to the results from prior analyses such as the homework.

A closer look at the students erroneous proof path solutions reveals that the two major difficulties are first that the students reverse the order of statements and reasons in the given proof problem. Most of the students who produced only a partially correct proof path solution, drew the arrows in the way they would arrange the statements and reasons in the SR-method. A students solution supporting this observation is shown in Figure 7 below.

The circled areas indicate erroneous reconstructions. For example, in area I the arrows are not drawn from the premises (=statements) to the rules (=reason), but to the conclusion first. In fact, looking closely at the generation of a proof path, a third part has been included in the proof - the conclusion of applying rules to premises. So every proof step in the proof path method consists of three parts. This is different from the statement-reason method where two parts are written out and the third is used to be the premise for the next step.

The second major difficulty is that the students do not keep the sequence of steps in the right order, although the sequence of proof steps is the same as in the statement-reason method. Nevertheless, it is probably a result of the lack of familiarity with the proof path method. Circled area II and III give an example of how proof steps can be set in the wrong order. Instead of drawing the arrows from all three premises to the rule SSS, an arrow was drawn from the premises to the conclusion of applying the rule SSS.

The evaluation answers after the task indicate that the students think that the SR method is still the preferred one, though 2/3 of the students doing the proof path think the proof is understandable in both ways. This answer is somehow in contradiction with the performance of the students. It might address a different issue that although the students still do not master this new method, they consider this method to be worthwhile approach.

Transportability of the System

The usual evaluation of ITS concentrates on two aspects: the internal and external evaluation. The external evaluation focuses on the different aspects of educational effectiveness. We think that an additional aspect, the transportability of the ITS, is not taken into account in this traditional evaluation approach. The next section describes how the Geometry Tutor and the current curriculum matched each other.

The matching of the GT and the curriculum

One of the authors' claim was '*because it does not do lesson presentation and because new problems are easy to generate, the tutor can be easily adjusted to fit any curriculum*' (Boyle 1987, p.1). If this claim is true then the students should not have any problems working with the GT. In fact, the way proofs were taught in the textbook differed in form and content from the way the proofs were done in the GT. In one problem the GT (Figure 9) required three additional steps before arriving at the same conclusion as the way presented in the textbook (Figure 8).

Most students (86%) thought that this was a fairly difficult problem for them due to the additional steps required. The students were in a similar situation with another proof problem. Comparing the results of the GT session and the homework, there was only a slight decrease in errors (from 86% to 80%) and the number of incorrect solutions speaks for itself.

Discussion

What are the results of this study? One aspect concerned the instructional effectiveness; we tried to answer the question whether the GT could teach the students geometry proof skills in the proof path method. We are not in a position to answer this question definitely since our results are ambiguous and we still did not cover the whole field to be tested. If the student had to apply outside of the GT the new proof method such as in the reconstruction situation, the performance was poor compared to the performance with the familiar SR method. A closer look at the errors revealed that most of them were due to the students confusion between the methods. Additionally, the GT required the students to be more explicit in their proof sequence. The percentage of perceived difficulties in this situation is very high. A comparison between the problems perceived at the actual interaction with the GT (86%) and the Homework 2 of this same task (80%), shows that there is no decrease. Apparently, it would take the student much more time to learn to integrate these additional proof steps. This could be seen as an indicator that the transition between the methods is not as smooth for students as one would think.

As for the transportability of the system, we can positively say that nearly all of the students enjoyed using the Geometry Tutor and would recommend it for other students. When the students were confronted with a difficult problem, the majority of them chose to do it with the GT. But we should not forget that at the end of the workshop about half of the students did not think that the GT was better to use to do proofs. Contrast these observations with a difficult integration into the classroom. There was a mismatch in at least two places (one reported in section 4.3.1) between the way the proof was done in the Geometry tutor and the method explained in the textbook. With a greater number of problems in the GT (there were only 17) one might imagine a greater number of these kind of mismatches. On the whole, the 'paper-pencil' test of the Geometry Tutor was not very successful. We understand the 'paper-pencil' test as a means to investigate what happens when the machinery is not available to the students.

At this point let us have a closer look at some of the GT features: the quality of help commands and the user interface. The GT got a good rating (AM=3.6, s=0.7) on a five-point scale for giving helpful hints on the proofs. We asked about the quality of the hints: 82% thought that the hints were easy to understand, 96% that they were appropriate for the problem and 87% that they provided enough help. Another question concerned to what extent the Tutor helped them to feel more comfortable with doing proof, the students answered that they felt quite comfortable (AM=3.5, s=1.2). It will be difficult to explain the students' contradictory answers. Although half of the students thought that using the GT was not better in doing proofs, the majority of them qualify the help provided by the GT as good. One explanation might be that the unfamiliar proof method overweights the students general impressions. Another aspect, the temporary 'rigidity' of the GT, appeared quite often during the actual interaction with the GT. Consider the following situation: a student decides to apply in the final phase of the proof the AAS rule to his/her givens. After the Inference dialog box appears on the screen, the student changes his/her mind and wants to apply the rule SSS (which in the described proof situation was possible since the student had all the necessary givens). But the student will be unable to leave the inference because the GT *thinks* that the student is on the right path. This is true but it neglects the possibility that in a given situation the user might decide that another proof step might be more appropriate. The described situation puts the user in a situation where s/he experiences helpnessless. Nothing can be done to go one proof step back. Several students experienced this situation at different steps in the proof.

In follow-up study (Kafai, in prep.) the Geometry Tutor will be used to explore alternative proof solutions. The students will be invited to reason about disadvantages and advantages of chosen solutions. We will investigate whether this helps the students to acquire better proof path skills, but also if it improves their reasoning skills. Additionally we will have a closer look at the tutorial intervention of the GT. Therefore we will record the actual interactions of the student with the GT and analyze the effect of the tutor's help. Results are forthcoming.

References

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, *228*, 456-467.

Anderson, J. R., Boyle, C. F., & Yost, G. (1985) The Geometry Tutor. *Proceedings of the International Joint Conference on Artificial Intelligence*, Los Angelos, pp.1-7.

Baker, E,. Feifer, R. Aschbacher, P., Bradley, C., & Herman, J. (1985). *Intelligent computer-assisted instruction (ICAI) study*. Los Angelos, CA: Center for the Study of Evaluation, University of California.

Bloom, B. S. (1984) The 2 Sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, vol. 13, pp. 4-16.

Boyle, C. F. (1987) *Geometry Proof Tutor User's Manual*. Carnegie-Mellon University, Pittsburgh, PA.

Bonar, J. G., & Weil, W. (1985) An informal programming language. Paper presented at the meeting *Expert Systems in Government*, October, Washington, D.C.

Brown, J. S., Burton, R. R., & de Kleer, J. (1982) Pedagogical natural language and knowledge engineering techniques in SOPHIE I, II and III. In D. Sleeman, & J. S. Brown (Ed.), *Intelligent Tutoring Systems* (pp. 227 - 282), London: Academic Press.

Burton, R. R. (1982) Diagnosing bugs in a simple procedural skill. In D. Sleeman, & J. S. Brown (Ed.), *Intelligent Tutoring Systems*, London: Academic Press.

Burton, R. R., & Brown, J. S. (1982) An investigation of computer coaching for informal learning activities. In D. D.Sleeman, & J. S. Brown (Ed.), *Intelligent Tutoring Systems* (pp. 79-98) London: Academic Press.

Dede, C. (1986) *A review and synthesis of recent research in ICAI*. International Journal of Man-Machine Studies, vol. 24, pp. 329 - 353.

Johnson, W. L., & Soloway, E. M. (1984) PROUST: knowledge-based program debugging. *Proceedings of the Seventh International Software Engineering Conference*, Orlando, Florida, pp. 369-380.

Kafai, Y. B.(1988) *The introduction of intelligent tutoring systems into the classroom setting: a case study of the Geometry Tutor*. Research Report #61, Cognition and Programming Project, Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor.

Lewis, M.'W., Milson, R., & Anderson, J. R. (1987). The TEACHER'S APPRENTICE: Designing an Intelligent Authoring System for High School Mathematics, in Kearsley, G. P. (Ed.), *Artificial Intelligence & Instruction*, Reading, MA: Addison-Wesley Pub. Co., pp. 269 - 301.

Littman, D. & Soloway, E. (1987) Evaluating intelligent tutoring systems: The cognitive science perspective, in J. Jeff Richardson, M. Polson (Eds.), *Intelligent Tutoring Systems*, Hillsdale, NJ: Erlsbaum.