BUILDING AN INTELLIGENT TUTORING SYSTEM: SOME GUIDELINES FROM A STUDY OF HUMAN TUTORS

Deborah K. Galdes and Philip J. Smith
Department of Industrial and Systems Engineering
The Ohio State University
Columbus, OH 43210

One reason that intelligent tutoring systems (ITSs) are rarely found outside of the research lab has to do with the guidelines available to developers of these systems. First, some of these guidelines are stated as general, abstract goals such as "adapt to the student." What ITS developers need, however, are specific strategies and techniques which can be implemented in an ITS to accomplish those goals. Second, not all of the guidelines have an empirical basis.

One solution to both of these problems is to study human tutors. This paper demonstrates this approach, and discusses an empirical study of human tutors which was conducted to address these issues. Specifically, it discusses

1) the knowledge acquisition method which we designed to capture the appropriate empirical data,
2) how we used this method to study human tutors and students in the medical problem-solving domain of immunohematology (blood banking),
3) several guidelines which appeared to drive the tutors' behavior (e.g., "limit the number of interrupts to the student"), and
4) specific tutoring strategies which can be incorporated into an ITS to make its behavior follow these guidelines.

INTRODUCTION

A promising technology available to designers of computer-based training systems is "intelligent tutoring systems" (ITSs). Yet, although ITSs have been around for over a decade, only a few are being used outside of the research lab (Dews, 1989).

One reason for this lack of use involves the guidelines available to developers of these systems. First, some of these guidelines are stated as abstract, general tutoring goals (Carroll & McKendree, 1987) such as "Minimize working memory load." (Anderson, Boyle, Farrell & Reiser, 1987). When developers sit down to <u>build</u> an ITS, however, these general guidelines/goals often leave them with the question of "But how and when do I <u>do</u> that?"

Second, not all of these guidelines have an empirical basis. As an example, the ITS guidelines developed for the WEST system were based on the researchers' own personal philosophy of coaching (Burton & Brown, 1979). This philosophy states that students should learn on their own as much as possible, and that human tutors interrupt far too often. Examples of guidelines from the WEST system include the following:

- Principle 5: "Do not tutor on two consecutive moves, no matter what."
- Principle 10: "If the student asks for help, provide several levels of hints."

We cannot rely on intuitively-based guidelines, though, because intuitions and "folk psychology" (Moran, 1981) can be contradicted by facts. One example is "Principle 5" stated above. This guideline is fine for a game situation such as WEST where the student is learning arithmetic skills, but what about a system that coaches students in a life- and death domain such as medical diagnosis? In medical diagnosis, a student might make two consecutive wrong moves with respect to generating hypotheses, collecting confirming and disconfirming evidence, and/or making inferences from the data available. Although all errors might not require immediate feedback, there is certainly a possibility that <u>not</u> tutoring on two consecutive moves might leave the student with a misconception that will someday put a patient in a life-threatening situation. Thus, without empirical data to support "Principle 5," it is difficult to determine if this guideline would or would not be appropriate to a medical domain.

One solution to both of these problems is to study human tutors. Studying human tutors will both

1) give us an empirical basis for our theories of tutoring and
2) provide insights for *how to implement* these guidelines.

This paper demonstrates this approach, and discusses an empirical study of human tutors which we conducted to address these issues. Specifically, this paper

1) describes an empirical study of human tutors who used the "case method" (Clancey, 1979) to tutor students in the medical problem-solving domain of immunohematology (blood banking),
2) presents several "guidelines" which appeared to drive the tutor's behavior, and
3) demonstrates how we can use the empirical data to define "tutoring strategies" which can then be incorporated into an ITS to make its behavior follow these guidelines.

The tutoring strategies that we identified are currently being implemented in an ITS which tutors students in this same domain of immunohematology.

METHODS

We viewed the problem of discovering a tutor's goals ("guidelines") and strategies as a knowledge acquisition task (Kidd, 1988). Both the data collection and analyses methods are summarized below and discussed in detail in Galdes (1990).

The Knowledge Acquisition Method

We conducted several pilot studies to investigate the usefulness of different knowledge acquisition techniques for our specific problem. The results of these pilot studies led to the following final design:

1) The tutor and student were in separate rooms with limited visual communication over a video link and controlled audio communication over an audio link. This setup differs from previous studies of human tutors which all used computer terminals as the communication link when the tutor and student were in separate rooms.

2) The tutor and student shared visuals over the video link. The cameras used to support this link were positioned to give the tutor and student a "bird's-eye view" of the other's hands and worksheets, and nothing else (i.e., not the other person's face). The tutor always had visual access to what the student was currently working on. The student only received visual information from the tutor when the tutor had specific drawings or diagrams to show the student. (This rarely occurred.)

3) The default situation was that the tutor was watching the student solve the problem via the video link and the audio link was disconnected. When either the student or the tutor wanted to initiate a conversation, they would press a button and the audio link would be connected. At this time, the tutor and student could talk freely to each other. When the conversation was finished, the audio link was disconnected. Whenever this link was disconnected, the tutors and students were instructed to "think outloud."

Our methodology demonstrated several important advantages. First, this methodology reduced the amount of nonverbal communication, yet retained face validity of the data by allowing tutors and students to carry on natural tutoring dialogues rather than typing at terminals. Second, it allowed us to, at times, collect verbal protocols from the tutors and students, giving us more insight into their cognitive processing. Third, tutors and students solved the cases using a familiar paper-and-pencil approach. This meant that we did not have to build a human-computer interface for solving cases in this domain before we knew what capabilities tutors and student required of such a system.

Selecting Cases

We used five cases in our study. The cases were structured so that the students would have been taught all of the concepts necessary to solve them. The tutor's job was to help students pull all of this knowledge together when solving actual cases.

Subjects

Our subjects included two tutors and ten students. Both tutors were practicing experts in the domain of immunohematology and had experience tutoring various levels of students and trainees in this domain. The students who participated in our study were all in the process of completing their first course in blood banking. Both the tutors and students received extensive training on various aspects of the task which they were not initially familiar with.

Tutoring Sessions

Tutors and students were not given any specific instructions on how to behave other than that the student would solve cases and the tutor's goal was to make sure that students could solve problems like these on their own when they had completed the tutoring session. Both the tutor and student could initiate conversations during the tutoring sessions. During these sessions, each student solved the same five cases. These sessions lasted from 2-3 hours per student.

In the case method, the student plays the role of the practitioner and begins each problem with the patient's case history. He is then expected to ask for further information and lab test results that might be relevant. We provided a "lab" for our students to get the necessary additional information. When students had finished with a case, they called the tutor to say, "I'm done."

The Domain of Immunohematology (Blood Banking)

The primary reason for selecting the domain of blood banking is that it exhibits many important features of medical diagnosis in general. This similarity increases the likelihood that our results will generalize to other medical domain.s The most important similarities include the following:

1) Solving problems in this domain focuses on generating hypotheses and collecting confirming and disconfirming evidence for testing these hypotheses.
2) Hypotheses can be generated, confirmed, and disconfirmed by requesting additional data in the form of additional laboratory test results.
3) There is an underlying qualitative model of the system being diagnosed (Clancey, 1986) which practitioners can use when solving problems in this domain.

Data Analysis Procedures

After collecting the necessary empirical data, our task was to analyze it for "tutoring expertise." The first step of the analysis consisted of creating hard-copy transcripts from the audio and video tapes. These transcripts included the following data:
The data which we collected consisted of

1) verbal protocols which were collected while the tutor was deciding whether or not to interrupt the student,
2) tutor-student dialogues which occurred after the tutor interrupted the student or the student requested help, and
3) the students' actions as seen in the tutor's view of the world.

The second step was to select one of the two tutors to model in detail. Our intent when collecting data on two tutors was to model one tutor in detail and use the other tutor for a later comparison. We felt justified in this approach because we can expect different experts to use different strategies to achieve the same result (Simon, 1980; Kail & Bisanz, 1982). Thus, building a generic model of both tutors would not necessarily represent a true model of expertise.

A preliminary analysis of the data demonstrated that the tutor's behavior partially revolved around the two important decisions of

- deciding whether or not to discuss a specific topic in the domain, and
- if necessary, deciding an appropriate time for initiating the discussion of that topic.

So, our third step was to mark all of the topics which the tutor seemed to consider discussing with the student, and describe all of the events associated with each of these topics. We termed each topic and its associated events an "incident." The associated events consisted of the student's action(s) at the time the topic arose, what else the student was doing at the time, what the tutor was verbalizing, and how this topic was eventually resolved (i.e., when it was dropped, discussed or reactivated as a slightly different topic). In all, we found 207 incidents for five students on four cases.

The fourth step of the analysis was to

1) determine the factors which seemed to influence the tutor's decisions of how to resolve these topics, and
2) develop rules which relied on these factors and logically described the tutor's behavior in the separate incidents.

These rules are the basis of the tutoring model described in Galdes (1990).

The fifth step was to determine the goals and reasoning ("guidelines") behind these tutoring rules whenever possible. Some of these were generated from the current literature (e.g., "adapt to the student") and tested against our data, while others arose during the analysis of the tutor's verbal protocol (e.g., "don't assume that correct student behavior implies correct knowledge").

GUIDELINES AND TUTORING STRATEGIES

We found seven "guidelines" which appeared to drive the tutor's behavior. This section demonstrates typical results from this approach by discussing three of these guidelines and what appeared to be the tutor's corresponding strategies. The guidelines include the following:

1) Adapt to the student.
2) Limit the number of interrupts to the student.
3) Don't assume that correct behavior on the student's part implies correct knowledge.

The remaining guidelines are in Galdes (1990).

Note that although most ITS designers would agree that the above guidelines are reasonable, many would be left with the question of "But how and when should my ITS do that?" Note also that the list of tutoring strategies presented for each guideline is not meant to indicate a complete and closed set, but rather an example of the kinds of results we can expect from an empirical study of human tutors.

Adapt to the Student

We found several different strategies which our tutor used to adapt to a student. These included the following:

1) *Look at the current context and the discourse history before deciding if the student's actions represent an error.* Moral: What is an error for one student is not necessarily an error for another student.

2) *Look at the cause of a definite error before deciding how to handle it.* That is, was it caused by a "slip" rather than erroneous knowledge? Moral: The "category" of error (e.g., missing goal vs. erroneous procedure) isn't the only influence on how to handle an error.

3) *Give the student extra time before interrupting him if he appears to not want any help.* Moral: Let the student's attitudes determine some of the parameters for tutoring.

4) *Set global constraints on the tutoring session as a whole based on different attributes of the student's personality/ability (e.g., shyness).* Moral: Don't feel obligated to use the same standard list of topics for each student. For example, discuss extra topics to get a shy student to talk more.

5) *In specific cases, adapt to a student's erroneous procedure* by teaching procedures for error detection and error correction rather than trying to correct how the student carries out a standard procedure. Moral: Don't always correct the student's procedure by attempting to rebuild existing pieces of the student's domain knowledge.

Limit the Number of Interrupts to the Student

This guideline follows the philosophy of WEST - limit the number of times it interrupts the student (Burton & Brown, 1979). To adhere to this guideline, WEST uses strategies such as "Do not tutor on two consecutive moves, no matter what." and "Do not tutor before the student has a chance to discover the game for himself." (Burton & Brown, 1979).

Our tutor, however, used very different strategies. These included the following:

1) Use a "wait-and-see" strategy whenever possible for discussing errors which arise when you are not engaged in a discussion with the student.

2) If a dialogue segment is initiated, then be efficient and also discuss topics "on hold" (i.e., those where a wait-and-see strategy is being applied).

3) Don't interrupt the student for every action that cannot be explained. If the student is likely to return to the correct solution path in the near future, don't interrupt.

4) If the student appears generally confused and he is likely to call soon, then wait. That is, let the student initiate the dialogue because the student's initiative might give some insight into the specific problem.

5) Do preventive maintenance. For example, if it is unlikely that the student knows what the next step should be, then ask about this next step at the end of a discussion about the current step. That is, occasionally probe for misconceptions rather than always waiting for them to arise from the student's actions.

Don't Assume that Correct Behavior Implies Correct Knowledge

A third guideline which the tutor followed was that she did not assume that correct behavior on the student's part implied that the student understood what he was doing. This guideline was explicitly stated in the tutor's verbal protocol.

To handle these "invisible" errors, the tutor had an entire category of topics which appeared to arise from her generic model of what students at this level did/didn't know. These were errors which were not normally evident from a student's actions. These topics included asking the student about the following types of potential errors:

1) Knowing the goal behind a correct action.
2) Knowing why the absence of an action was correct in some cases.
3) Being able to predict the outcome of a lab result.

Note that for each of the above, the tutor did not ask about every possible instance. For example, she did not ask a student to state the goals behind all of his actions. To decide which to discuss, the tutor appeared to use a generic model of what students at this level were likely to know/not know.

CONCLUSIONS

The tutoring guidelines and strategies described in this paper demonstrate two important points. First, what ITS designers really need when they sit down to build a system are the tutoring strategies which suggest *how to implement* the guidelines found in the literature. Second, studying human tutors is a feasible method for generating empirically-based guidelines and the tutoring strategies necessary for implementing these guidelines.

ACKNOWLEDGMENTS

This work was supported by the National Heart, Lung, and Blood Institute Grant #HL-38776 awarded to Jack Smith, Jr. in the Pathology Department at Ohio State University.

REFERENCES

Anderson, J.R., Boyle, C.F., Farrell, R. & Reiser, B.J. (1987). Cognitive principles in the design of computer tutors. From *Modelling Cognition*, P. Morris (Ed.), John Wiley & Sons.

Burton, R.R. & Brown, J.S. (1979). An investigation of computer coaching for informal learning activities. *International Journal of Man-Machine Studies, 11*, 5-24.

Carroll, J.M. & McKendree, J. (1987). Interface design issues for advice-giving expert systems. *Communications of the ACM, 30.*

Clancey, W.J. (1979). Tutoring rules for guiding a case method dialogue. *International Journal of Man-Machine Studies, 11*, 25-49.

Clancey, W.J. (1986). Qualitative student models. *Annual Review of Computer Science, 1*, 381-450

Dews, S. (1989). Developing an ITS in a corporate setting. *Proceedings of the Human Factors Society 33rd Annual Meeting*, pp. 1339-1342.

Galdes, D.K. (1990). *An Empirical Study of Human Tutors: The Implications for Intelligent Tutoring Systems*, Doctoral Dissertation, The Ohio State University.

Kail, R.V. Jr. & Bisanz, J. (1982). Cognitive strategies. From *Handbook of Research Methods in Human Memory and Cognition*, C.R. Puff (Ed.), Academic Press.

Kidd, A. (Ed.) (1988). *Knowledge Acquisition for Expert Systems: A Practical Handbook*, Plenum Press.

Moran, T.P. (1981). An applied psychology of the user. *Computer Surveys, 13*, 1-10.

Simon, H.A. (1980). Cognitive science: The newest science of the artificial. *Cognitive Science, 4*, 33-46.