American Educational Research Journal
Spring 1998, Vol. 35, No. 1, pp. 65–99

How Tutors Model Students: A Study of Personal Constructs in Adaptive Tutoring

Sharon J. Derry and Michael K. Potts
University of Wisconsin–Madison

Studies of tutoring are producing rich descriptions of tutorial dialogue but have not identified constructs that tutors use to classify and discriminate among students for the purpose of adapting tutoring to student differences. This study investigated five experienced tutors' personal constructs about students tutored over a significant period of time. Several tutoring settings and domains were represented. Constructs used by tutors to discriminate among tutees were identified with repertory grid interviews and interpreted with the aid of cluster analysis. All tutors judged and classified students in terms of two underlying dimensions that were similarly defined, though not exactly alike, across tutors: motivation and intellectual ability. Tutors' personal constructs and tutorial decisions informed by those constructs are reported, and implications for programming computer-based tutors are discussed.

SHARON J. DERRY is a Professor, Department of Educational Psychology, Education Sciences Building, 1025 W. Johnson St., Madison, WI 53706. Her specialization is cognitive psychology applied to education.

MICHAEL K. POTTS is a School Psychologist, Indian Prairie School District 204, Box 693, Oswego, IL 60543. His specialization is school psychology.

Human tutoring provided on a one-to-one basis has been credited as the most effective form of instruction (Bloom, 1984; Cohen, Kulik, & Kulik, 1982). It is not surprising, then, that efforts to isolate and describe the actions of expert tutors and the unique interactions that take place between tutor and tutee have intensified in recent years. Cognitive researchers, particularly those interested in developing intelligent computer learning environments, are seeking models that specify how good human tutors behave and think during tutorial interaction.

Through the use of verbal protocol and interaction analysis, techniques that focus on dialogue between tutor and tutee, researchers are examining in detail the nature of human tutorial interaction. Conversational analysis of tutoring sessions, focusing on such issues as how tutors frame tutoring sessions, ask questions, aid students in problem solving, and react to student errors, has been the dominant methodology (e.g., Fox, 1993; Graesser & Person, 1994; Hume, Michael, Rovick, & Evens, 1995; Lepper, Woolverton, Mumme, & Gurtner, 1993; McArthur, Stasz, & Zmuidzinas, 1990; Merrill, Reiser, Merrill, & Landes, 1995; Person, Kreuz, Zwaan, & Graesser, 1995; Putnam, 1987). Such studies have produced rich, detailed descriptions of tutoring in action in specific contexts. Before presenting our study, which employed a different methodology, we overview findings of this research.

One of the earliest efforts (Putnam, 1987) found little evidence to suggest that human tutors build detailed cognitive models as a basis for understanding student performance and adapting tutoring strategy. This finding was interesting because error diagnosis, knowledge modeling, and identifying bugs in student thinking were major foci of intelligent tutoring research of the time (e.g., Sleeman & Brown, 1982; Van Lehn, 1988). Putnam's results suggested that human tutors (much like expert teachers in studies by Leinhardt and Greeno, 1986) are guided more by curricula and agendas than by detailed student models. Thus, Putnam's tutors tended to be more procedural than reactive over the course of tutoring sessions. Responses to errors were corrections designed to steer students back to the more profitable path of the original script.

These findings can be compared to those of McArthur et al. (1990), who conducted a detailed study of algebra tutoring of adults. Like Putnam, McArthur et al. failed to find evidence that errors triggered detailed diagnosis of specific student knowledge states. However, tutors set up and conducted performance assessments and devoted significant effort to remediation. Tutors called on numerous techniques and complex scripts, and the preponderance of these were used for remediation and task management, where task management included introducing and framing of problems, and breaking them down into manageable chunks.

Fox (1993) conducted a series of studies that examined effects on tutorial dialogue of factors, such as differences in academic domain. She also sought to describe the structure of tutorial dialogue and how tutors and students handled intervention and communicative repair. Fox found that tutoring was very complex and that students played a major role in negotiating the learning process that took place. Tutors used errors, student responses to queries, and features of interaction (e.g., the timing of student responses, the way in which a response was delivered) as diagnostic evidence for adapting their tutoring. Correcting student mistakes was a major focus of tutoring, although either the student or tutor was likely to initiate correction and tutors often withheld correction until the student had time to initiate. Correcting by the tutor in conceptual domains was often by questioning.

A study by Merrill et al. (1995) characterized tutor assistance in a highly procedural domain, LISP programming. Their study provided a fine-grained analysis of human tutoring over extended periods of problem solving. They observed that different forms of tutor response to students' errors resulted from tutors' estimates of the learning consequence of errors. For example, to avoid the dangers of floundering that can result from a syntactic error, tutors were more likely to directly inform a student how to repair the error. Semantic errors in a student's program, however, were more likely to be followed by the tutor's indicating the general location of the programming mistake. If necessary, student and teacher would then engage in a collaborative negotiation about where the student's reasoning was faulty and how they might reconstruct a more useful solution strategy.

Responses to student errors also figured prominently in a study by Hume et al. (1995), who found that hinting was a major strategy used by expert tutors to help medical students learn to reason causally and qualitatively about a complex physiological system. Working in a computer-based, problem-solving environment in which communications between tutor and tutee were filtered through a computer interface, tutors responded to student problem-solving errors by giving hints that either reminded students of known information or (if necessary) conveyed needed information. Tutors also used complex hinting strategies to guide students through longer chains of reasoning. In situations where students did not respond, hinting was quickly abandoned in favor of more direct approaches.

A different picture of tutoring has been painted by Graesser and Person (e.g., Graesser & Person, 1994; Person et al., 1995), who studied interactions of unskilled student tutors working with less advanced students in mathematics and statistics. They reported the prevalent, though flexible, use of a 5-step tutoring script: (a) Tutor asks a question; (b) student answers; (c) tutor gives feedback on answer; (d) student and tutor collaborate to improve quality of answer; and (e) tutor assesses the student's understanding of the answer. However, unskilled tutors frequently gave inappropriate feedback. For example, polite affirmations were provided at times when negative or corrective feedback would have been beneficial. Also, unskilled tutors seldom paused for students to correct their own mistakes. They immediately spliced in information when an error was detected. Surprisingly, students made academic gains even with unskilled tutoring.

Although studies mentioned thus far did not explicitly identify motivation as a major factor in tutoring, numerous tutoring techniques reported were similar to motivational tutoring techniques found by Lepper et al. (1993) in their studies of teachers who tutored children. Lepper et al. suggested that human tutors monitor and respond to affective, as well as cognitive, states of students. Their tutors attended to student motivation and adapted their tutoring with strategies such as attributing difficulties to tasks rather than ability and varying challenge, curiosity, and control within tutoring sessions.

The picture of human tutoring that emerges from research is not easily summarized. As pointed out by Hume et al. (1995), tutoring appears to vary widely and is greatly influenced by such factors as pedagogical skill of the tutor, type of subject being tutored, subject expertise of the tutor and student, and the institutional context in which the tutoring takes place. Yet it is possible to draw a number of qualified generalizations:

1. Skilled human tutors often structure their tutoring sessions around a basic preconceived agenda, which may be adapted to meet student needs.
2. Expert tutors possess and call on various tutoring strategies, scripts, and routines for adapting instruction and guiding action; unskilled tutors generally follow a more limited question-asking script.
3. Skilled tutors spend time on task management—framing tasks with helpful explanation, breaking down complex problems into manageable steps.
4. In natural settings with skilled tutors, students negotiate with tutors and may even be responsible for setting the tutoring agenda and for helping to characterize the tutoring session and the dialogue.
5. Responding to errors is an important part of tutoring, and skilled and unskilled tutors differ in their error-response patterns. In skilled tutoring, error correction is made in various ways, at different levels of specificity, and may even be initiated by students. Skilled tutors often pause for students to see and repair their errors. Unskilled tutors tend to respond immediately, directly, and often inappropriately.
6. In skilled tutoring, errors are treated as learning opportunities, although not all types of errors present the same opportunities for pursuing instructional goals. Thus, some errors may be pursued in depth, while others are corrected immediately and directly.
7. Although some degree of cognitive modeling does take place during tutoring, even skilled human tutors do not appear to construct elaborate, detailed models of students' curriculum knowledge or lists of their buggy concepts to use as a basis for adapting tutoring.
8. From studies of tutorial responses, it is evident that skilled tutors are better than unskilled ones with respect to reasoning about problem solving from the student's point of view. Thus, skilled tutors must construct better situational models of students' problem solving than unskilled tutors do.

9. Tutors obtain diagnostic information from observing both *what* the student says and does and *how* (tone of voice, inflection, hesitancy, etc.) it is said or done.
10. Studies of tutoring children suggest that tutors attend and respond to both cognitive and affective/motivational student states; studies of adult tutoring have not interpreted tutoring behavior in motivational terms.

Individual Differences and Tutors' Personal Constructs

From a developer's point of view, what is missing from the picture above is information about what knowledge and which processes intervene between the observation of diagnostic evidence by the tutor on the one hand and the selection of tutoring techniques and styles on the other. Accepting that human tutors probably do not conduct bug-level diagnoses or construct complex and detailed models of students' knowledge or understandings as a basis for tutorial decision making, what knowledge do human tutors use to help them make tutoring decisions? An assumption on which we based this work is that, during tutoring, the expert tutor gathers evidence, forms relatively general impressions of the student, and then uses these impressions in deciding what kinds of tutoring might work best. Although decisions about what topic to remediate are based on performance assessments that help determine which curriculum objectives have or have not been mastered, decisions about the amount of interface control given over to the student, timing of interventions, and other elements of tutoring style (Lepper & Chabay, 1988) are likely to be based on knowledge about students and how they differ from one another.

We hypothesized that experience leads tutors to evolve constructs that they use to describe and discriminate among students on the basis of general categories relevant to tutoring. For example, consider the judgment, "Although this student is not very advanced, I think I will give her this exercise because she is motivated and likes to be challenged." Here, the hypothetical tutor is using general student-difference categories (level of motivation, general level of advancement) to justify the choice of exercise. Choices of this type are referred to as *global* decision making because they represent decision points that occur before or between intense tutorial dialogues. For example, deciding before a session what objectives will be pursued, how much control over the learning process will be ceded to the student, or what problem sets will be used in the session is a global-level decision (in which the student may or may not participate). This global decision can be contrasted with *local* tutorial decision making, which is more opportunistic. An example of local decision making would be deciding how to respond to a particular student error (e.g., whether or not to correct it). Like global decisions, such local decisions can also be informed by a tutor's personal constructs about students. For example, the tutor might reason, "This student seems frustrated, so I will correct this error directly." The construct being used to describe the student is frustration level. We hypothesized, then, that tutors possess constructs they use for describing tutees and that these help tutors build general student models that inform decision making at both local and global levels.

This study identified mental constructs used by five experienced tutors to help them characterize and tutor their students. To isolate these constructs, we employed a research method originally inspired by George Kelly's (1955) personal construct theory, which has been used frequently in expert systems development (e.g., Boose, 1985) but, to our knowledge, never before applied in intelligent tutoring research. Kelly proposed that individuals live and operate within specific worlds or domains (or communities of practice?) that can be broken down into the major components or elements that comprise them. An individual's perspective and understanding of that world can be described in terms of the constructs he or she uses to distinguish among and think about the world's elements. For example, if students are regarded as elements within an experienced tutor's world of tutoring, the ways in which students can be modeled by that tutor are determined by the constructs the tutor uses to describe and categorize them for the purpose of adaptive tutoring.

We combined Kelly's repertory grid interview methodology with statistical analysis to elicit a set of fundamental concepts that tutors used to describe and differentiate among tutees with whom they had recently worked and were very familiar. We also acquired information from each tutor regarding which constructs were most important in tutorial decision making and how they were used within a sample of tutoring sessions.

The following describes results of using this method to study two sets of skilled tutors. The first set comprised three tutors who tutored basic mathematics and worked with students in a computer-assisted laboratory setting. The second set comprised two tutors employed in a remedial college support program, one who tutored mathematics, and one who tutored language arts. A detailed description of the method is provided first. This is followed by results and a discussion of findings obtained from the computer-assisted tutors. Next, results and a discussion of findings for tutors in the college remedial environment are presented. A concluding discussion offers a summary interpretation of findings that focuses on similarities and differences among tutors and draws implications for the design of computer-based intelligent tutors.

Method

The Tutors

Five experienced tutors who practiced in a variety of educational contexts (i.e., across academic disciplines and instructional settings) were studied. Three of these five tutors were examined on their experiences tutoring college students using the interface of TiPS (Derry, Wortham, Webb & Jiang, 1996), a computer-based instructional system designed to improve basic mathematics and problem-solving ability. One of the computer-aided tutors worked with 12 student dyads, while one worked with 8 and the other with

9 individual students. These tutors were all members of the research team developing the instructional system and consequently were very familiar with the cognitive-theoretical foundations of the system development project, including the purpose and methods of this study. Using project researchers as model tutors in studies is becoming common practice on intelligent tutoring projects (e.g., Goodyear, 1992; Hume et al., 1995; Van Lehn, personal communication, July, 1995), and arguments have been proffered in favor of this procedure (Goodyear, 1992).

The computer environment used by project tutors consisted of a graphics interface that required students to learn five basic math concepts and the diagrams that represent them. Once acquired, the diagrams can be used by students to help them analyze and solve both simple (single-step) and complex (multistep) story problems on the system. Through the process of analyzing and solving problems on the system with tutorial guidance, students are supposed to hone their conceptual understanding of basic math and develop better problem-solving skills. Tutorial interaction was studied in the context of this interface because a practical purpose of this research was to develop a model of tutoring that could serve as the basis for designing the automated tutoring component of this and similar systems.

In addition, to gain information about the variability or stability of tutoring constructs across settings and subject domains, and also to draw comparisons with tutors working in natural, nonlaboratory environments, we studied two additional tutors not associated with the intelligent tutoring project. These participants were employed as tutors in a program designed to provide academic support for high-risk college students. One of these tutors helped 12 students with their math classes. The other tutor helped 7 students with writing assignments.

Data Collection Procedure

The data collection procedures were adapted from George Kelly's (1955) personal construct theory as well as from the knowledge acquisition techniques for expert systems development described by Boose (1985). Sessions with tutors consisted of two main phases, knowledge elicitation and rating grid.

Prior to their knowledge elicitation sessions, tutors were asked to prepare a list of tutees with whom they had recently worked and were extremely familiar. They were also advised that they should gather and bring available notes and records of tutoring sessions to the knowledge elicitation sessions and that they could use such records to help them recall during the session.

During knowledge elicitation, each tutor was presented with a series of triads, each triad consisting of three of the names from the list of tutees well known to that tutor. The names in each triad were determined by random selection with replacement. The ordering of triads and ordering of names within triads were also randomized. Triads were presented to and examined by the tutor one at a time. For each, the tutor was told to think of the most important attribute that distinguished the two most similar members of the presented triad from the third outlying member. The discriminating construct reported by the tutor was recorded along with what the tutor provided as an opposing construct (e.g., the construct *involved* was recorded along with what the tutor believed to be an opposing construct, *uninvolved*), creating the two ends of a bipolar scale which represented the range of description for each construct. This presentation of triads and identification of discriminating constructs continued until several constructs given by the tutor merely repeated previously reported constructs and no obvious new constructs were emerging. Statements and questions presented to the tutors are reported in the knowledge elicitation protocol included in the appendix.

Tutors were also asked to recall tutoring sessions as a basis for specifying *what* evidence they observed that enabled them to discriminate among students on the basis of each construct and to state *how* such discriminations influenced their tutoring of these students. To the extent that this technique attempted to focus tutors' retrospective recall on specific tutoring sessions and tutees and also provided tutors with various memory aids (notes and records) to prompt specific memories, we believe much of the generated data fall within the *many-to-one* category (one verbalization for many actions), as defined by verbal protocol theory (Ericsson & Simon, 1984). This theory supports the assumption that data obtained in this manner are a veridical representation of what actually happened in tutoring, albeit much less dense than what think-aloud methods would provide. However, tutors were also asked a few questions that required them to generalize and make inferences across tutoring sessions. For example, they were asked which constructs were most important overall for adapting their tutoring. For these types of questions, verbal protocol theory makes no predictions regarding how accurately answers reflect actual performance.

Once a representative list of tutor constructs and their polar opposites was elicited, a rating grid was developed so that the tutor could score all tutees on all of the bipolar scales based on the constructs elicited by that tutor. The rating grid was constructed by listing all tutees in a row at the top of the page and listing the bipolar scales in a column on the left of the page. Each student was then rated by the tutor on all construct scales. Not all constructs were evaluative, but those thought to possess a pole representing a more positive characterization of a student were placed so that the highest rating (5) was given for the most positive characterization, while the lowest number (1) represented a rating toward the negative end of the pole. For example, if the construct were *interest*, then the rating scale would be 5 = *very interested* and 1 = *very bored*, with the midpoint (3) representing *neither interested nor bored.*

Results

Tutors Working in a Computer-Based Laboratory Setting

Tutors each rated their tutees on their personal construct scales, and these ratings were transformed into *z* scores for input to SPSS (Norusis, 1993) for Windows. *Z* scores were computed based on within-tutor and within-scale means. Cluster analyses were then performed for each tutor. Cluster analysis is a statistical procedure that identifies conceptually homogeneous groups or clusters of cases based on values given for a set of variables.

Two types of cluster analyses were computed for each tutor. One analysis treated bipolar constructs as cases. Clusters of bipolar constructs represented the basic, underlying dimensions of judgment used to describe and rate tutees. The other analysis treated students as cases and thus created clusters representing groups of students who were similar in terms of how they were perceived by the tutor. Student clustering represented a search for naturally occurring student types, or generic student models.

In the following sections, findings of cluster analyses will be reported separately for each tutor. Identical analytic procedures were employed for all tutors, so some procedural details are given only once for the first tutor (Tutor JS).

Tutor JS

JS was a project researcher with more than 10 years of classroom teaching experience in college and high school and about 100 hours of experience as a tutor in both language and mathematics. Judging from years of formal graduate training and work experience, JS was the most knowledgeable of the five tutors with respect to cognitive instructional theory.

JS's tutees were eight undergraduate students (1 male, 7 female), previously unknown to JS, enrolled in a large psychology course at a major university. All had failed to achieve mastery on a word problem pretest and expressed interest in improving their problem-solving ability. Students were offered a choice of extra course credit or financial compensation for their participation. JS tutored seven of the eight students for about 5 hours each. The eighth student was available for only 4 hours of instruction. Most students came for 1-hour sessions, although two strong students were allowed to have 2-hour sessions.

In response to the questions, "What was your main purpose, and what were your objectives in the context of the tutoring you did with these students?" JS cited the following:

1. Learn/review basic concepts and the system notation for them.
2. Use system to solve 1-step problems of all types.
3. Use system to solve 2- or 3-step problems involving mixed concepts.
4. Learn strategy of working backward from the goal.
5. Write problems for 1- to 3-concept combinations.
6. Solve complex, multistep problems of mixed types.
7. Tutor other students or write explanatory worked examples for others.

JS's stated purpose was to move students as far as possible through this "curriculum agenda," on the "faith" that these activities would improve students' in-depth understanding of basic math. The agenda was not followed in strict order.

Cluster analyses. Based on the knowledge elicitation procedure, 28 bipolar constructs were identified by the tutor. These are listed in Table 1 as they appeared in the hierarchical cluster analysis of JS's constructs, shown in Figure 1. (The Arabic numeral associated with each construct indicates the order in which it was mentioned by the tutor during the interview.) Headings in Table 1 are our suggested names for clusters of constructs, based on our study of the tutor's detailed definitions.

As verified by Figure 1, JS's constructs clearly defined higher order

Table 1

Clusters of Tutoring Constructs Used by JS With Most Important Constructs Highlighted (in Italics) and Ranked (in Parentheses)

<u>Motivation</u>

12. Not curious / Curious
24. Easily bored / Interested
20. Does not enjoy TiPS / Enjoys working on TiPS
11. *Not interested in theory / Highly interested in theory (I)*
22. Closed, not talkative / Very talkative, open
25. Frustrated / Not at all frustrated
27. Criticized/disliked system / Liked system
13. Does not understand/accept ability limits / Understands/accepts ability limits

<u>Learning ability</u>

Intelligence/Aptitude

2. Exceptionally weak / Exceptionally strong
6. Difficult to teach / Easy to teach
8. *Not competent / Competent (II)*
10. Learns slowly / Learns fast
19. Does not reach advanced levels / Reaches advanced levels
7. Plodding / Quick
9. Struggles to learn / Learns with ease
28. Mastered only low-level skills / Mastered high-level skills
16. Indecisive / Decisive

Cognitive effort, concentration

5. Careless / Careful
18. *Poor concentration / Strong concentration (V)*
14. *Shallow thinker / Deep thinker (III)*
15. Ill at ease (distracted) / Confident
23. Gained little / Made strong gains
26. *Failed to master basic concepts / Mastered basic concepts (IV)*
21. Uncomfortable with human tutor / Comfortable with human tutor

<u>Atypicality</u>

3. Typical student / Atypical student
4. Average / Unusual
17. Has no opinions / Opinionated
1. Weak basic math background / Strong basic math background

categories. One category was defined by constructs used to discriminate among students on the basis of their affective, motivational states. The other characterized students in terms of their learning abilities and itself had a substructure: One subcluster pertained to general intelligence, the other to cognitive effort. A third, smaller main cluster emerged that appeared to measure the degree to which students were perceived as either typical or atypical, where atypical students tended to be rated as opinionated and reported a strong math background.

Responding to the question of which constructs most influenced tutoring, JS listed the following in order of importance: interest in math theory, general competence, depth of thinking, concept mastery, and level of concentration. These are shown highlighted with Roman numeral rankings in Table 1. Thus, four of this tutor's top five constructs were from the learning abilities cluster, but the most important discriminator of students was a motivational construct.

The second cluster analysis, performed to identify naturally occurring

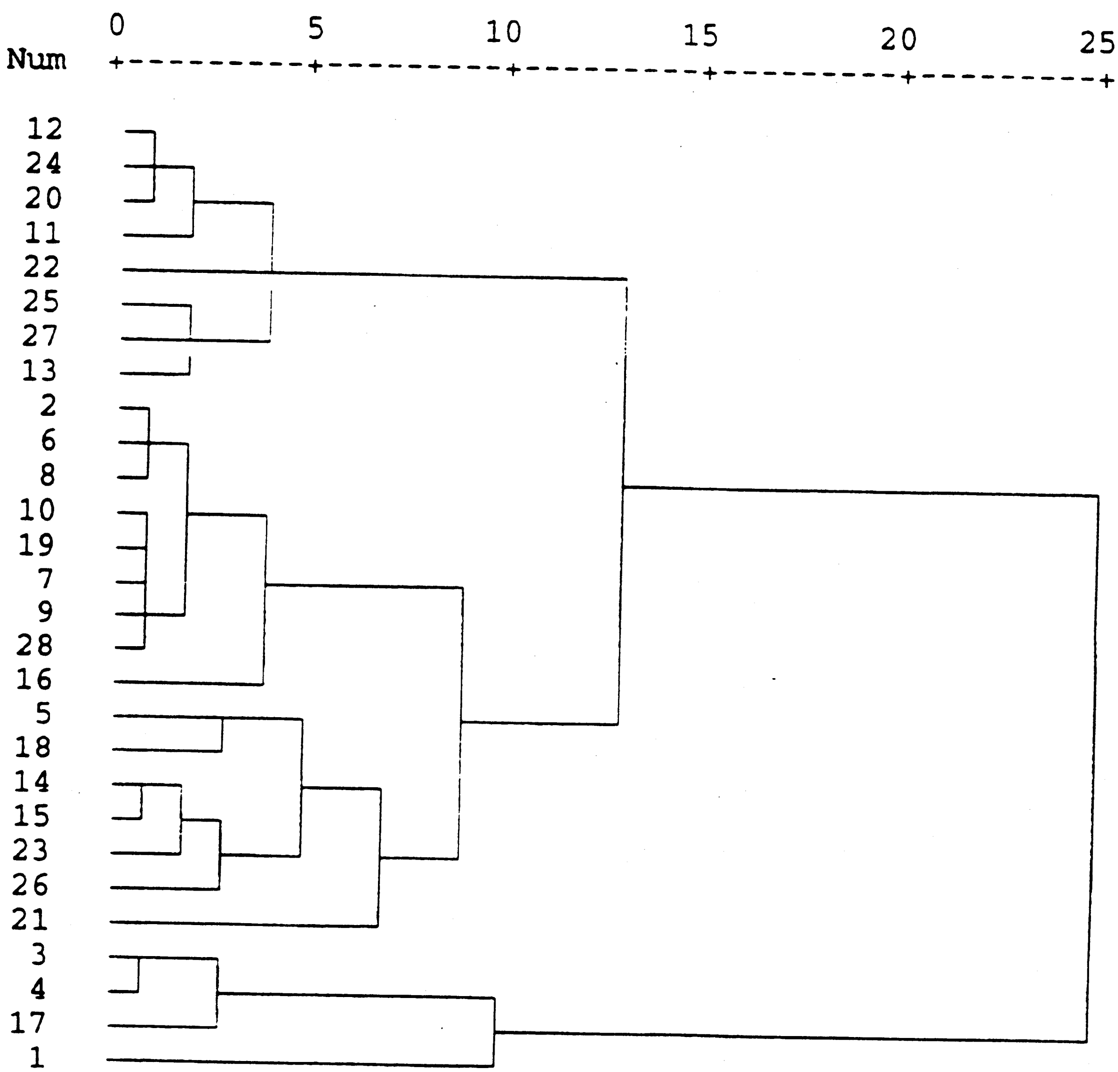

Figure 1. **Hierarchical cluster analysis for JS's constructs**

groupings of students, is shown in Figure 2. The cluster defined in Figure 2 by Students 5, 6, 7, and 8 grouped students who scored moderately on both ability and motivational scales and were called "typical students" by the tutor. Student 2 was not a member of any cluster and was the only student perceived as highly motivated but very low in ability. Two students (3 and 4) clustered together as being relatively high in both motivation and ability. Student 1, JS's "difficult student," was also not included in any cluster and was the only student JS perceived to be very low in both ability and motivation (although this student claimed to have excelled in high school math).

Figure 3 depicts how Tutor JS grouped students as *types* in terms of their relative positions on a grid representing the tutor's two main discriminating constructs (motivation and ability). To determine student positions on the grid in Figure 3, students' ratings on constructs associated with each of the two main superordinate constructs were averaged to produce an overall motivation and an overall ability rating for each student. These were then plotted on the grid, and student clusters were indicated by drawing circles around student groups. The large higher order Cluster A in Figure 3, grouped seven nonproblematic students together, discriminating them from the sole problematic student shown within Cluster B. The analysis further indicated that the nonproblematic group was composed of three subtypes: A.1—the typical students who were moderately bright and motivated; A.2—a single student who was highly motivated but had very poor skills; and A.3—those who were exceptionally high in both motivation and ability.

Tutor PM

Identical analyses were performed for Tutor PM, whose previous teaching experiences included working with university students on career development and in an introductory sociology course. PM had also worked for several years as a school psychologist and had taken graduate courses in cognitive instructional theory. PM's tutees included nine undergraduate students (7 female, 2 male) recruited from the same population as those of JS, who were also offered either extra course credit or financial compensation for participation. Individual sessions ranged from 60 minutes to 80

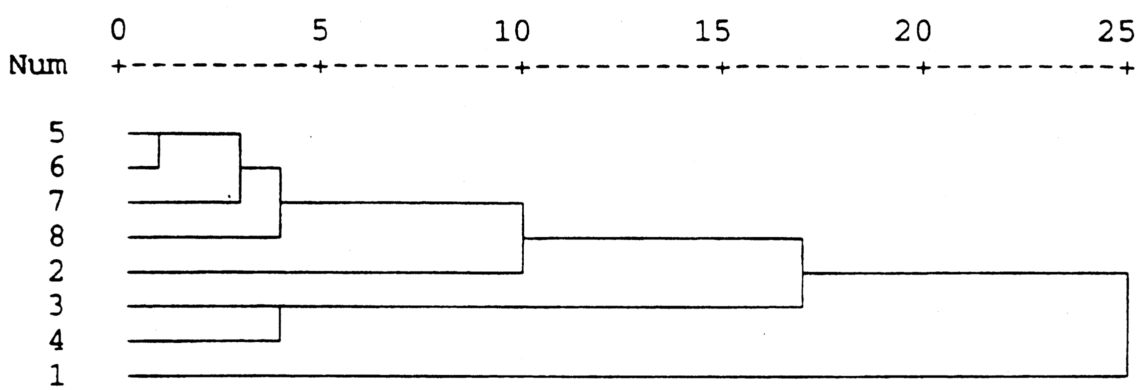

Figure 2. **Hierarchical cluster analysis for JS's students**

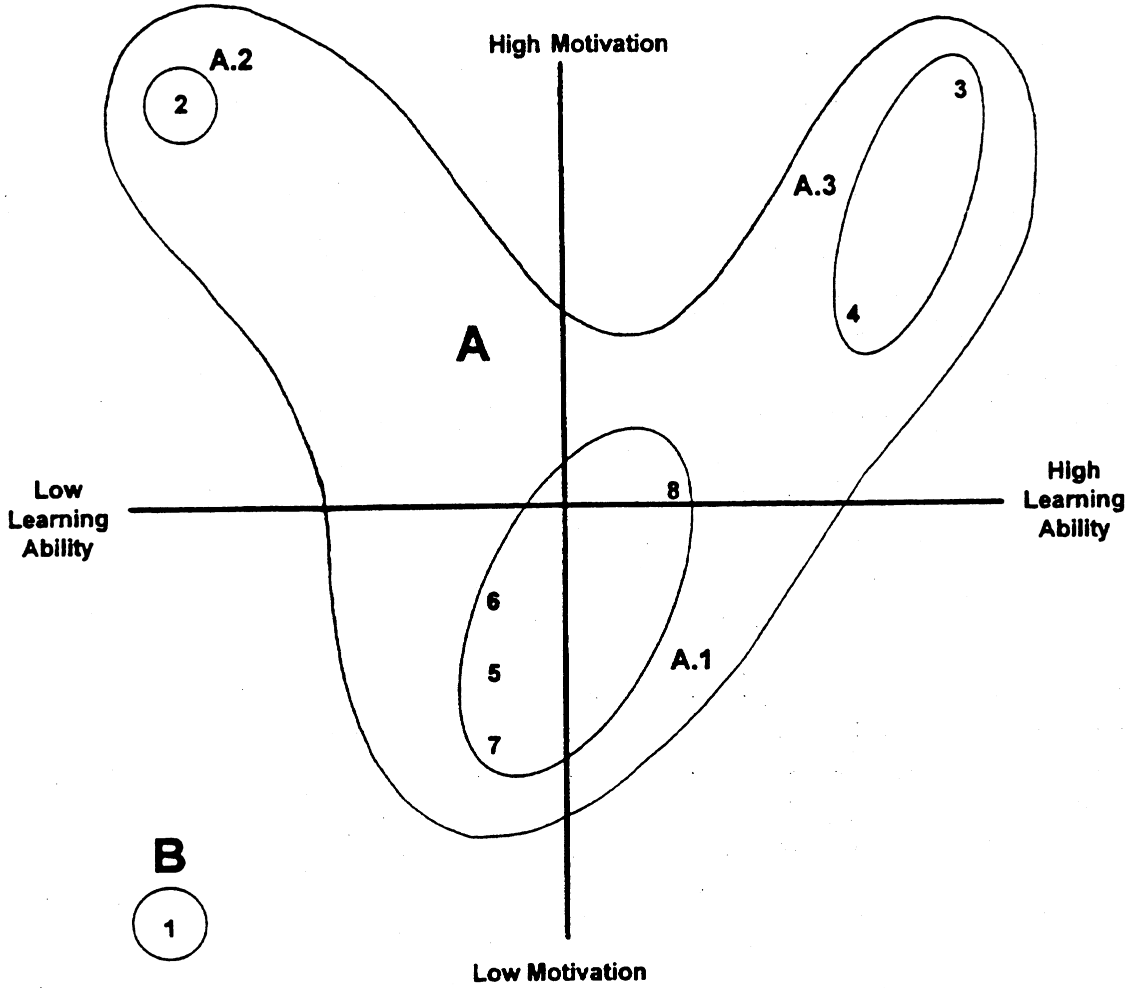

Figure 3. **Student clusters for computer-assisted Math Tutor JS**

minutes. One homework assignment was included after the first tutoring session.

PM provided the following statement of tutoring objectives:

> My main purpose . . . was to get them to progress as far as they could with the program in the time that we had together. In order to do this I wanted to establish a good rapport with each of the students so that they would also enjoy the social aspect of working with a tutor. An ongoing goal was to increase the students' independent and competent use of the program.

Cluster analyses. Based on the knowledge elicitation procedure, 17 bipolar constructs were identified by the tutor as important for describing and discriminating among tutees. These are listed in Table 2 as they were grouped by a cluster analysis of constructs.

PM's personal constructs clustered into two higher order categories. One was defined by constructs used to discriminate among students on the basis of their overall problem-solving competence on the computer system.

Table 2

Cluster of Tutoring Constructs Used by PM With Most Important Constructs Highlighted (in Italics) and Ranked (in Parentheses)

System competence

4. Not self-directed / Self-Directed
12. Does not select challenging problems / Selects challenging problems
9. Weak understanding of concepts and notations / Good understanding of concepts and notations (II)
13. Early trouble with function concept / Early competence with function concept
6. Low degree of confidence / High degree of confidence
10. Not sure of answers / Sure of answers
5. Needed support past first three sessions / Independent past first three sessions
3. Not quick in learning software / Quickly learned to manipulate software

Affective response to system

1. Low tolerance of frustration / Good tolerance of frustration (I)
11. Not interested in graphics use / Interested in graphics use (V)
8. Did not seem to enjoy tutoring sessions / Enjoyed tutoring sessions
14. Wanted to quit early / Wanted to solve one more problem
15. Nonreflective / Reflective (III)
17. Wanted tutor to drive interface / Wanted to drive interface

Prior computer-related experiences

16. Low prior computer experience / High prior computer experience

Approach to labeling diagrams

2. Not concerned about label completeness / Meticulous about labeling (IV)

Degree of computer integration into problem solving

7. Preferred to write out problems first / Worked problems directly on computer

The other characterized students in terms of their affective/motivational reactions. Three independent and unique constructs that did not cluster with others also were identified. These pertained to prior computer experience and to specific ways in which students used the computer. The five constructs ranked by PM as the most relevant for tutoring are shown in Table 2. Three of these, including the top one, were associated with the affective/motivational dimension; one was associated with the competence dimension, and one was an independent construct that clustered alone.

The second cluster analysis, showing how PM grouped students as types, is depicted in Figure 4. As shown, all except one student (No. 5) were grouped in a broad cluster of students whose affective/motivational reactions to the instruction ranged from moderately to extremely positive. Within this cluster, there were several subclusters roughly indicative of different competence levels. Student 5 was not a member of any cluster and was the only difficult student perceived as very low in both ability and affective response.

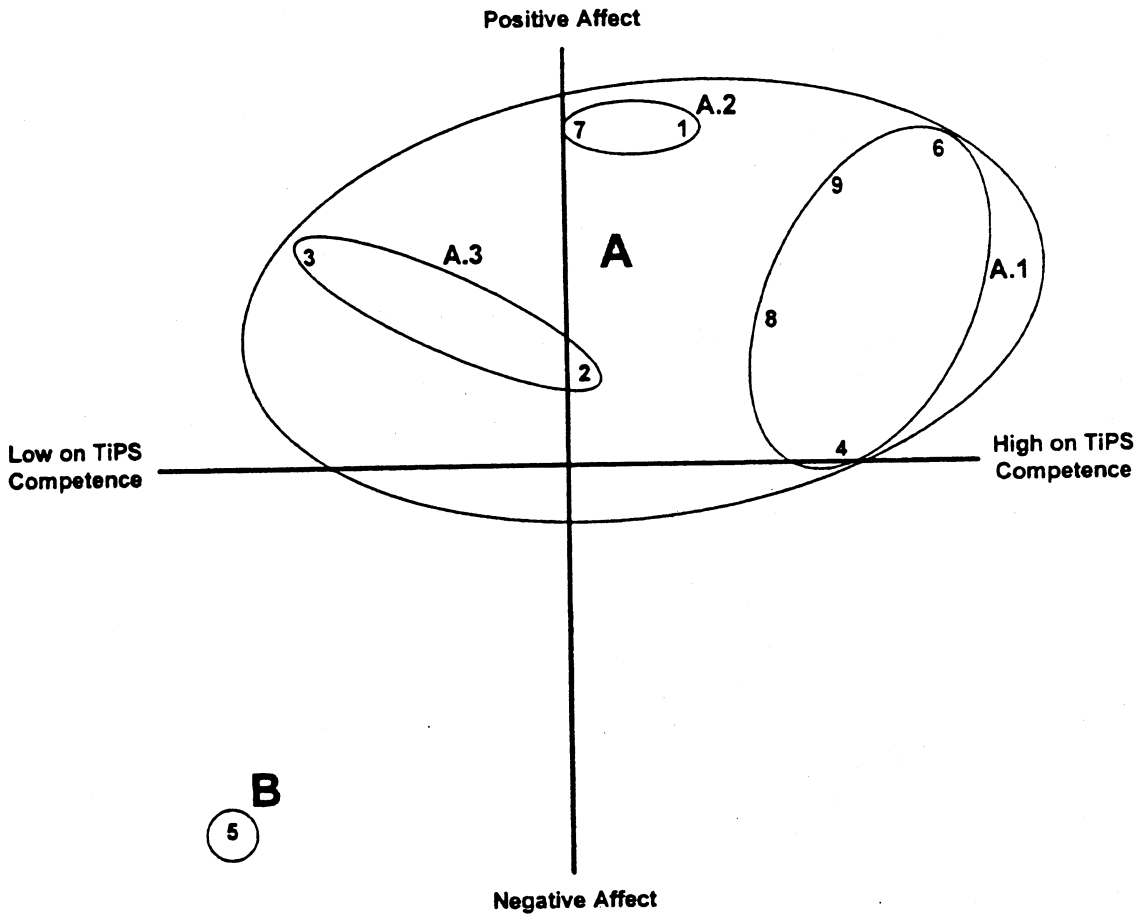

Figure 4. **Student clusters for computer-assisted Math Tutor PM**

Tutor KT

KT's background included several years as a teacher of computer skills at a small college. As a graduate student specializing in cognitive science, KT worked as a paid statistics tutor for other graduate students. KT tutored 24 university students (17 female, 7 male) in dyads on mathematics word problems as part of another study using the TiPS computer system. KT's tutees were recruited from the same population as those of JS and PM, but KT's volunteers had been more extensively tested to ascertain that they lacked the specific skills being tutored. With each pair, KT conducted five 75-minute training sessions.

In describing tutoring objectives, KT reported feeling restricted, because tutoring was associated with a study and had to follow a strict protocol. The protocol involved tutoring pairs of students to mastery on two skill sets. The first skill set included learning the interface and its five graphic objects, representing different math concepts. The second skill set was three problem-solving strategies. Before moving to the second skill set, students had to achieve mastery on the first. Before taking the final test, students had to demonstrate mastery of the target strategies. A worked-example instructional method was employed, where to-be-learned concepts and skills were demonstrated for students by the tutor, followed by having them work practice problems with tutorial assistance. In sum, KT's sessions were much more scripted than those of PM and JS.

Cluster analyses. Fifteen bipolar constructs were used by KT to describe and discriminate among student pairs. They are listed in Table 3 as they were grouped by cluster analysis, which produced three main clusters plus a single independent construct. The first two clusters were similar to the general ability and affective/motivational clusters appearing for Tutors JS and PM and were named Pair Competence and Pair Engagement, using labels supplied by KT. The main conceptual deviation from other tutors' motivation clusters was that the scale—*sustained motivation, slipping motivation*—appeared in KT's competence, rather than his engagement, cluster. However, from interviews, it was evident that, for KT, this scale meant presence or absence of negative verbalizations related to students' perceived competence, such as "I can't do this." The third cluster, as well as an additional, independent scale, represented a collection of constructs whereby the tutor discriminated pairs on the basis of their performance on particular problems or skills that were specific to KT's study. Although these extra groupings emerged, all of the top five constructs named by KT as most

Table 3

Clusters of Tutoring Constructs Used by KT With Most Important Constructs Highlighted (in Italics) and Ranked (in Parentheses)

<u>Pair competence</u>

1. Struggle / No struggle (I)
2. Worst leader / Best leader
7. 1- & 2-step problems poorly done / All 1- & 2-step problems correct
14. Slipping motivation / Sustained motivation (II)
11. Slow / Fast (V)

<u>Pair engagement</u>

4. Uninvolved / Involved (IV)
9. Doesn't follow directions / Follows directions
13. Nonreflective / Reflective
12. No mentoring of partner / Mentoring of partner (III)

<u>Situation-Specific issues</u>

8. Difficulty with metric conversion problems / No difficulty with metric conversion problems
10. Hard time formulating backward strategy problems / Able to do backward strategy problems
3. Switching strategy during problem solving / Choosing & completing chosen strategy
5. Missed relevant information on pizza problem / Used all relevant information on pizza problem
15. Tried illegal interface actions / Didn't try illegal interface actions
6. Solves politics problem before working on computer / Solves politics problem ON computer

important for tutoring were from Pair Competence and Pair Engagement. Table 3 shows rankings for KT's top five constructs. Three, including the top one, were from the competence cluster, while two were from the engagement cluster.

The second cluster analysis, performed to identify groupings of student pairs, is illustrated by Figure 5. As shown, two clusters of student pairs were discerned, the largest (labeled A on Figure 5) including all pairs scoring above the midrange on Pair Engagement and the smaller (B) comprising the two remaining problem pairs who scored very low on both competence and engagement. Within the first higher order group, there were several subgroups of pairs. The largest group (A.1 on Figure 5) was composed of students who rated relatively high on both competence and engagement. The second cluster (A.2) was for a single pair that was highly engaged but relatively low on competence. A third, loosely connected, cluster (A.3) included two pairs, both moderately engaged, although their competence differed.

Evidence Observed by Computer-Assisted Tutors and Their Tutorial Adaptations

As developers of a computer-based tutoring system, we were interested in

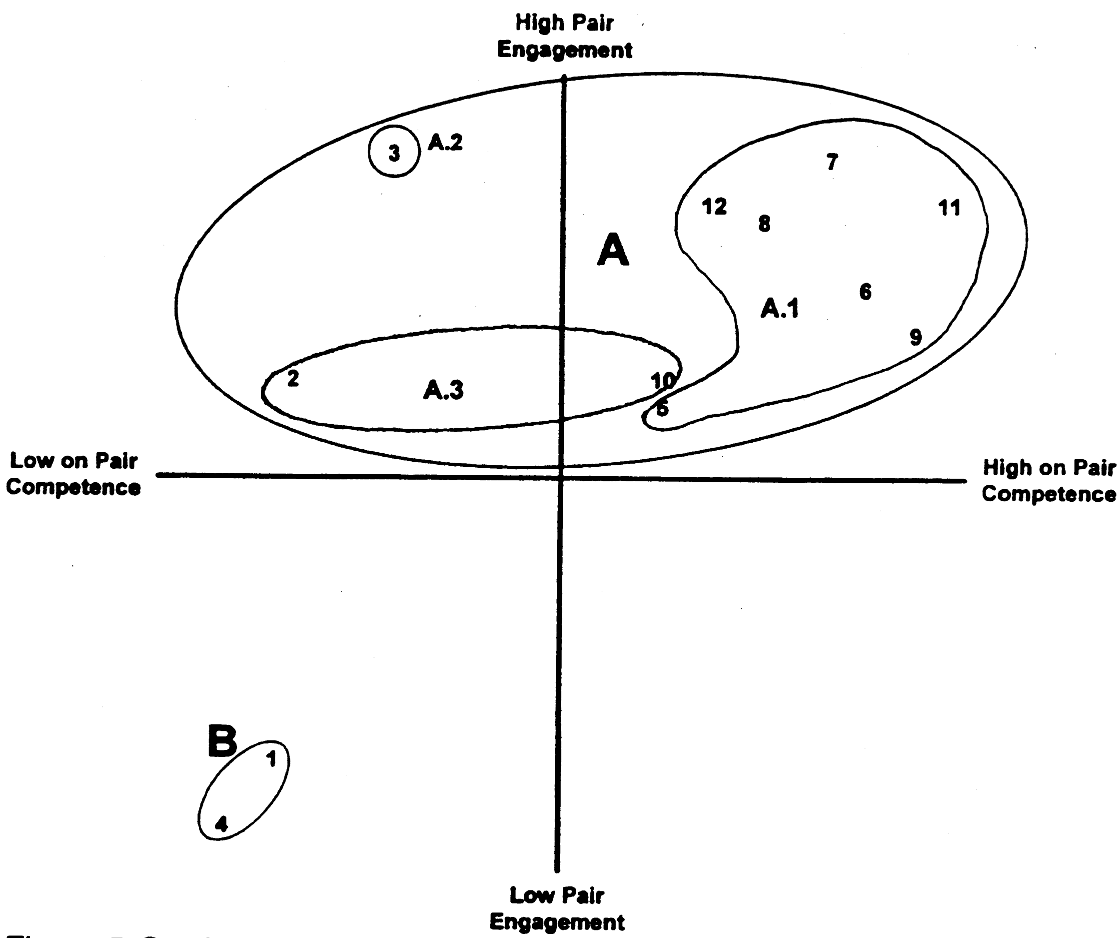

Figure 5. **Student pair clusters for computer-assisted Math Tutor KT**

knowing what specific behavioral indicators were observed by human tutors to help them judge the motivational and ability levels of tutees and what specific tutorial adjustments were made in response to estimates of motivation and competence. Tables 4 and 5 provide a listing of the reported behavioral evidence used by the three computer-assisted tutors to determine students' ability or competence (Table 4) and their motivational/affective states (Table 5). Although not evident in Tables 4 and 5, in a few cases, tutors reported combining these behavioral indicators to draw complex inferences relevant to tutoring. For example, students seen as low ability and also as indecisive and hesitant were viewed by JS as fearful and lacking in confidence, whereas those seen as low ability and decisive were viewed as careless and nonreflective.

Tables 4 and 5 also list what tutors said were adaptive responses to their judgments about students' competence and motivation. These lists represent reported tutorial adaptations only, not complete lists of conversational moves that might characterize each tutoring strategy in its entirety. Tutors who did not mention hinting, for example, did not report adjusting their hinting strategy on the basis individual student differences. Failure to mention hinting in this context would not indicate the absence of hinting in tutoring.

Asterisks in Tables 4 and 5 indicate the tutors' indicators and responses we believed could not be detected or programmed with reasonable ease into the instructional system that we are developing. Notably, only a few indicators and responses should be difficult or impossible to accomplish with our computer-based technology.

Discussion of Adaptive Tutoring in the Computer-Based Laboratory Environment

All computer-assisted math tutors reported monitoring students' general competence and affective/motivational levels during tutoring and altering tutoring strategy accordingly. For example, and as shown in Table 4, JS reported using competency assessments to determine the pace of interface instruction, the sophistication of theoretical discussions, the degree of student control permitted, the extent of review and verbal guidance provided, and the tendency to wait for students to correct errors. As shown in Table 5, motivational assessments reportedly influenced such things as decisions to talk about theory at all, variety in instructional routine, willingness to hear critical comments from students, degree of humor and conversational informality in dialogue, and whether students' personal interests would be evaluated and problems adapted to those interests. These kinds of adaptations stand in contrast to the idea of adaptive tutoring based on specific knowledge models, which would be indicated if tutors had frequently discriminated among students or altered their tutoring strategy in response to specific knowledge differences, which they did not. But while tutors did not frequently discriminate among students on the basis of their specific knowledge differences, the speed with which students accomplished particular instructional objectives did appear to inform general competency assessments. For example, JS indicated that students who did not master basic concepts quickly were viewed as less capable and consequently were moved more slowly with more verbal guidance through a curriculum agenda. However, JS later relied on many other types of evidence to modify early competence assessments.

All three computer-assisted tutors constructed similar general student models, each model representing a categorical type of student that reportedly received particular versions of tutoring. All discriminated *Problem Students* from *Other Students*. Problem Students were characterized by low motivation and ability. Other Student models had several submodels. All tutors identified *Exceptional Students*, who were high in ability and motivation, and *Eager Students,* who were high in motivation but not ability. One tutor (JS) also discriminated a *Typical Student* category, comprising students who were moderate in both motivation and ability. Notably, no computer-assisted tutor in this study saw any student as both capable and unmotivated.

Tutors differed from one another in several ways. JS, the most experienced and educationally advanced (with respect to instructional theory) tutor, differed from the other two by producing the greatest number of discriminating constructs and a relatively complex personal construct space that was characterized by a two-dimensional (concentration and intelligence) view of student competence. This tutor's space of student groupings was comparatively complex, because it was characterized by the greatest number of tutorial actions. Finally, JS differed strikingly from the other tutors on a characteristic that might best be described as tutoring style. In sum, JS appeared to be the least empathic (e.g., Lepper & Chabay, 1988) of the three tutors, a characterization best explained by the following comparison with Tutor PM.

In contrast to JS, who focused largely on student competence, Tutor PM reportedly considered and reacted to students' affective states. PM reported that evidence of student negativity was immediately countered by acknowledgment and encouragement. No comparable statements were reported by Tutor JS. Possibly due to PM's empathetic stance, only one student was viewed by PM as having a motivational problem. By contrast, JS rated four students as moderately to very unmotivated.

The interviews of PM and JS reflected distinctively different philosophical orientations regarding beliefs about what tutoring should accomplish. JS's stated purpose was to move students through a "curriculum agenda." With weaker students, this was reportedly accomplished by providing direct verbal instructions. Stronger students were given little or no assistance with problem tasks, although JS reported drawing these students into discussions of theory. By contrast, PM reportedly emphasized the goals of independence, understanding, self-evaluation, and enjoyment for all students, fostering progress toward goals by questioning students in ways that forced them to evaluate their performance and by strategically withdrawing tutor support even if it was desired. Except for interface issues, little direct

Table 4

Evidence Used by Three Computer-Assisted Math Tutors to Judge Tutees' Learning Ability and Tutors' Responses to High and Low Learning Ability

	Behavioral evidence observed by tutors	
	High learning ability	Low learning ability
Tutor JS	Quickly learned interface	Learned interface slowly
	Eventually mastered system graphics	Never mastered system graphics
	Quickly solved early problems	Slowly solved early problems
	Made few conceptual errors	Made frequent conceptual errors
	Spontaneously corrected many errors	Seldom corrected errors spontaneously
	Made frequent productive queries	Seldom made productive queries
	Took charge of own work	Depended on tutor
	Gave extended attention without prompting	Prompting required to maintain attention
	*Regularly thought aloud about performance	*Seldom thought aloud about performance
	Typed carefully, infrequent input mistakes	Typed impulsively, made input mistakes
	Student was alert	Student seemed sleepy or fatigued
	Seldom used trial-and-error strategies	Often used trial-and-error strategies
	Made no comments about tutor presence	Said that tutor watching was distracting
Tutor PM	Demonstrated independent problem solving	Requested or required much tutor involvement
	Selected challenging problems when given choice	Selected easier problems when given choice
	Used interface independently after first three sessions	Misunderstood interface past three sessions
	High ratio of correct to incorrect solutions	Low ratio of correct to incorrect solutions
	Skillfully used function graphic	Unable to use function graphic
	*Used system to conduct deep problem analysis	*Preoccupation with surface (interface and/or labeling) concerns
	Was confident in final solutions	Was not confident with final solutions
	Easily recalled rules of interface use	Was unable to recall rules of interface use

Tutor KT	Minimal effort expended to solve problems correctly *Pair had strong leader Mastery of 1- and 2-step problems Did not complain about ability with system/task Demonstrated high performance on required skills Completed many problems in time limit	Problem solving required great effort *Pair had no strong leader Continual difficulty with 1- and 2-step problems Spoke negatively about ability with system/task Gave poor overall performance on required skills Completed few problems in time limit
	Adaptive tutorial responses	
Tutor JS	Interface instruction given quickly Held sophisticated discussion of theory Student given choice of topics and activities Little verbal instruction or encouragement given For easy objectives; little instructional variety given Little or no review provided at beginning of sessions Allowed student to correct errors	Significant time spent on interface discussion Held little or no theoretical discussion Tutor chose most topics and activities Much verbal instruction and encouragement given For easy objectives, a variety of routines used Review provided at beginning of sessions Pointed out most errors
Tutor PM	*Engaged student in discussion using system vocabulary Reinforced correct use of interface Reinforced correct reflections about current state of problem solving Reinforced independent problem solving Reinforced independent self-evaluation	Encouraged adoption of system vocabulary Recalled and restated rules of interface use Encouraged more verbal reflections on current state of problem solving Explicitly noted students' overreliance on tutor Encouraged more active self-evaluation process Switched roles, having student "instruct" tutor
Tutor KT	Gave few encouraging expressions Moved to more difficult problems	Gave frequent encouraging expressions Offered easier problem on same topic, either problem with fewer steps or one from more intuitive setting in which tutee had more background

*Indicates observation or behavior that would be very difficult for computer.

Table 5

Evidence Used by Three Computer-Assisted Math Tutors to Judge Tutees' Motivational States and Tutors' Responses to High and Low Motivation

	Behavioral evidence observed by tutors	
	High motivation	**Low motivation**
Tutor JS	Asked questions, explored Made comments / suggestions •Was sociable, friendly Oriented / attended to task Adapted to system: Praised system, used with ease	Did not question or explore Made no suggestions or comments •Was unsociable, not friendly Drifted, was distracted Didn't adapt to system: Blamed system, had difficulty
Tutor PM	Resolved frustrations Solved extra problems Was reflective, thoughtful Questioned Expressed interest	Remained frustrated Did no extra work Was impulsive, quick Asked few questions Expressed lack of interest
Tutor KT	Cooperated (followed directions) Was reflective, thoughtful •Gave partner help Displayed positive affect	Was uncooperative (did not closely follow script) Did not think things through •Ignored partner Displayed negative affect
	Adaptive tutorial responses	
Tutor JS	Provided elaborate theory Encouraged critical input Interacted with humor, lightness Ignored interface	Avoided theory Discouraged criticism Avoided corny humor, straight-forward Explained, justified interface Varied activity Personalized lesson and problems Shortened lessons
Tutor PM	Reinforced success Reinforced interest •Answered all questions	Encouraged taking time Reinforced success Acknowledged frustrations •Encouraged and reinforced thinking aloud
Tutor KT		•Initiated questioning strategy designed to draw in reluctant students

•Indicates observation or activity that would be very difficult for computer.

instruction was given by PM. Rather, content was conveyed by discussions that employed the conceptual language of mathematics. Perhaps due to PM's clinical background, the tutoring as described was noticeably therapeutic in style, while that of JS was either didactic (for less capable students) or laissez-faire (for strong ones). Asterisks in Tables 4 and 5 indicate that the tutoring style of PM would be relatively difficult to program.

Studies of Tutors in Natural Environments

To help contrast and compare the tutoring that occurs in computer-based laboratory environments to that which occurs in natural settings not involving technology, two tutors employed by a remedial support program for special admissions college students were studied using the same analysis techniques that were employed with the computer-assisted laboratory tutors. MJ tutored 12 students in mathematics, and MR tutored 7 students in communication arts.

Tutor MJ

MJ conceptualized the tutor's role as one of assessing what students already knew about the material being covered in their math classes, finding out what questions students had about class material, and then providing supplemental instruction and assignments designed to remediate deficits related to class requirements.

The knowledge elicitation interview drew on MJ's experience with 12 students. A total of 14 constructs were identified by MJ as important for discriminating among students for instructional decision making. These are listed in Table 6 as they were grouped by cluster analysis. Two fairly distinct clusters, plus several relatively independent scales, emerged. The first cluster, characterized by constructs referring to active involvement, persistence, and attitude, was labeled *Motivation*. The next cluster, defined by achievement, ability, and academic confidence, was labeled *Academic Aptitude*. As shown in Table 6, four of MJ's top five constructs for tutoring, including the top one, were from the motivation cluster. In sum, MJ was similar to other math tutors in that MJ discriminated among tutees on the basis of their motivation and ability. For MJ, the more important of these two concerns was motivation.

The cluster analysis of MJ's students is illustrated in Figure 6. Cluster A represents tutees who were low or relatively low on both motivation and academic aptitude. The remaining tutees were all part of a higher order cluster representing students who were positively motivated, either moderately or strongly. Within the motivated group, there were several subclusters representing different aptitude levels. Examination of Figures 3–6 revealed that clustering patterns for MJ were similar to those of the computer-assisted math tutors (especially PM and KT), although MJ's problem-student group was larger.

Table 6

Clusters of Tutoring Constructs Used by MJ With Important Constructs Highlighted (in Italics) and Ranked (in Parentheses)

Motivation

8. Likes to work in groups / Wants-demands individual attention
10. Passive, involuntary participation / Active, voluntary participation (II)
4. Lackadaisical / Structured
2. Negative attitude / Positive attitude
13. No desire for help / Reaching out for help (III)
1. Lacked persistence / Persistent (I)
7. Motivated enough to get by / Extremely, highly motivated (V)
5. Average work habits / Superior work habits

Realistic self-appraisal

11. Didn't see limits of own ability / Saw limits of own ability

Academic aptitude

3. Low achievement / High achievement
14. Low ability / High ability (IV)
9. Lack of confidence / confident in academic ability, achievement

Social tendencies (independent scales)

6. Quiet / Bubbly
12. Disruptive in groups / Not disruptive in groups

Tutor MR

MR was a graduate student, who was also a former English teacher, and had substantial tutoring experience in language arts. MR tutored seven university students, including three who were also tutored by MJ. The purpose of tutoring was to assist students with writing and public speaking assignments. MR met individually with tutees for three to four sessions per class project. Each session lasted approximately 50 minutes, and each tutee was tutored for four different projects during one semester.

MR provided the following description of tutoring objectives.

> My primary goal is always to get them to work independently, to recognize the strength of their own ideas. I work to pull ideas out of them, to organize their ideas, and keep the tutoring session very clearly focused on them and their thinking. . . . When I step in to be directive it is usually directing them toward more independence. . . . It is really having them actively think, trying to push them to think more critically than they thought before.

Seventeen constructs used to describe and categorize students were elicited and are shown in Table 7 as they were grouped by cluster analysis. The first of MR's major clusters was clearly an affective/motivational cluster, similar to those elicited from math tutors in many respects but also contain-

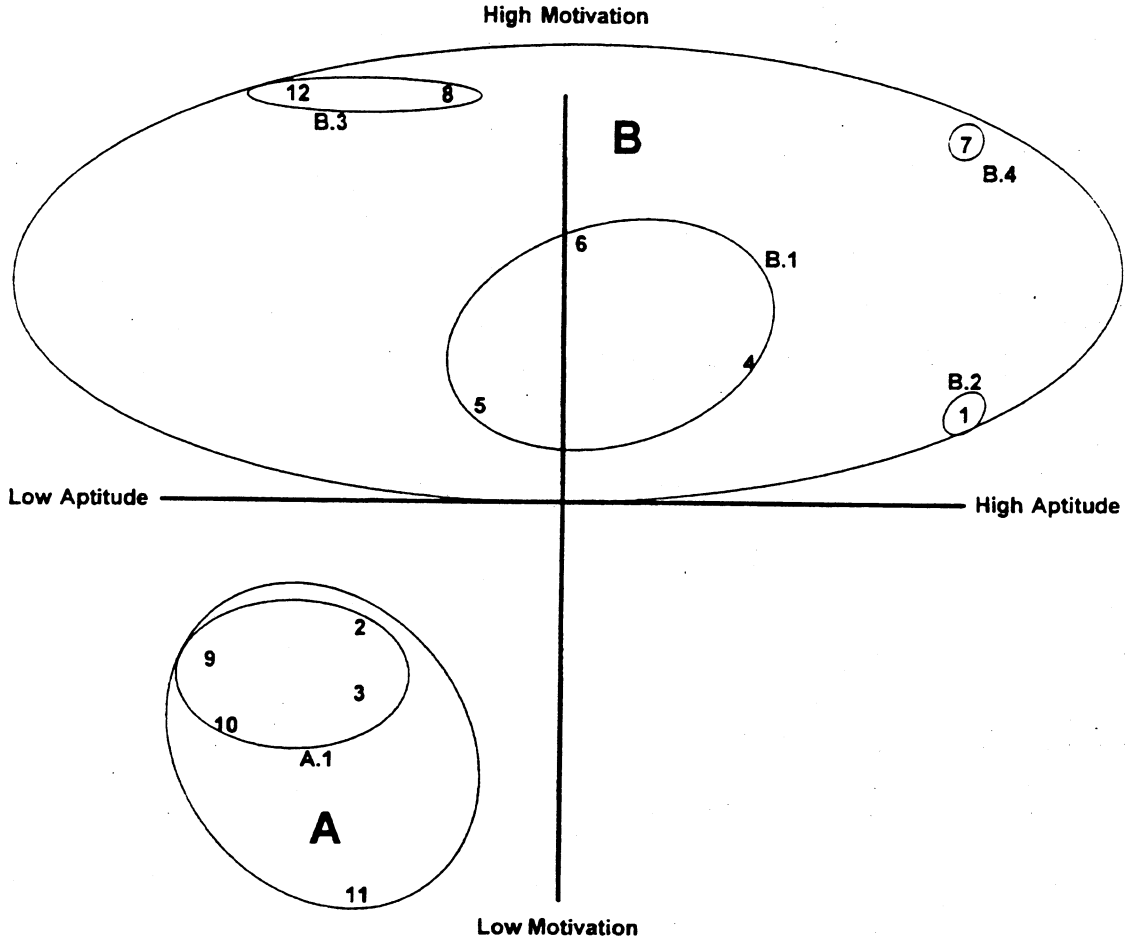

Figure 6. **Student clusters for Math Tutor MJ**

ing items related to what might be called *writer's block*—procrastination with writing assignments. The second major cluster of personal constructs pertained to learning ability and was easily relatable to similar clusters obtained from math tutors in that it subsumed such ideas as students' brightness, level of engagement in tasks, and speed of performance. MR ranked three motivational and two achievement constructs as most important for tutorial decision making. Notably, the top three were from the motivational cluster.

In the cluster analysis of students (see Figure 7), Cluster A grouped students who rated high or moderately high on learning ability but relatively low on motivation. Students in Cluster B were consistently rated positively on all constructs of the motivation cluster, although Student 4 was perceived as having much lower ability and thus clustered only loosely with Students 3 and 6. In contrast to all math tutors, MR perceived no student as having both low ability and low motivation. Most students were viewed as capable of thinking and writing well, though some chose not to.

Evidence Observed by Tutors in Natural Environments and Their Tutorial Adaptations

Like computer-based tutors in laboratory environments, tutors in natural

Table 7

Clusters of Tutoring Constructs Used by MR With Most Important Constructs Highlighted (in Italics) and Ranked (in Parentheses)

Motivation

4. Resists producing text / Does not resist producing text
9. Stays in oral discussion phase / Moves from oral to written expression (III)
5. Procrastinates, doesn't use time well / Paces assignments well
10. Didn't want instruction on drafts / Seeks instruction on drafts
2. Resistant / Engaged (II)
11. Off-task / On-task (I)
17. Motivation based on immediate assignment / Motivation based on long-term goal
15. Doesn't seek help during idea development / Seeks help during idea development

Ethnic awareness

8. Personal biases interfere with learning material / Objective approach to learning

Personality

13. Defensive / Not defensive
7. Low self-awareness of ethnicity / High self-awareness of ethnicity

Learning ability

6. Slow / Bright
12. Doesn't synthesize course materials / Synthesizes course materials (IV)
3. Needs leading (pushing) / Self-Directed (V)
14. Not personal with tutor / Personal with tutor
16. Plodding, slow / Last-minute, fast
1. Learning disabled / Not learning disabled

environments observed tutees' behaviors and used these observations as evidence for classifying tutees on the basis of motivation and ability. Some evidence observed was very similar to that observed by tutors in computer-based laboratory environments. As examples, MJ and MR judged motivation on the basis of questions asked by tutees (like JS and PM), willingness to tackle extra work (like PM), and negative comments (like all other tutors). However, in accordance with the broader social context of their tutoring, different types of evidence were mentioned. For example, both MJ and MR judged students' motivational levels on the basis of missed appointments. Similarly, the tutorial responses available to MR and MJ were also expanded by the natural-environment context. For example, the natural-environment tutors took on a more active parental role in calling and inquiring about students' progress and general well being.

Because MJ worked with math and MR tutored language arts, the strong appearance of motivational and ability clusters for both suggests there is stability of global tutoring constructs across academic disciplines as well as across tutors. However, a striking difference that distinguished MR from others was MR's perspective that almost all students were academically capable. Though one unusual student was motivated but not capable due

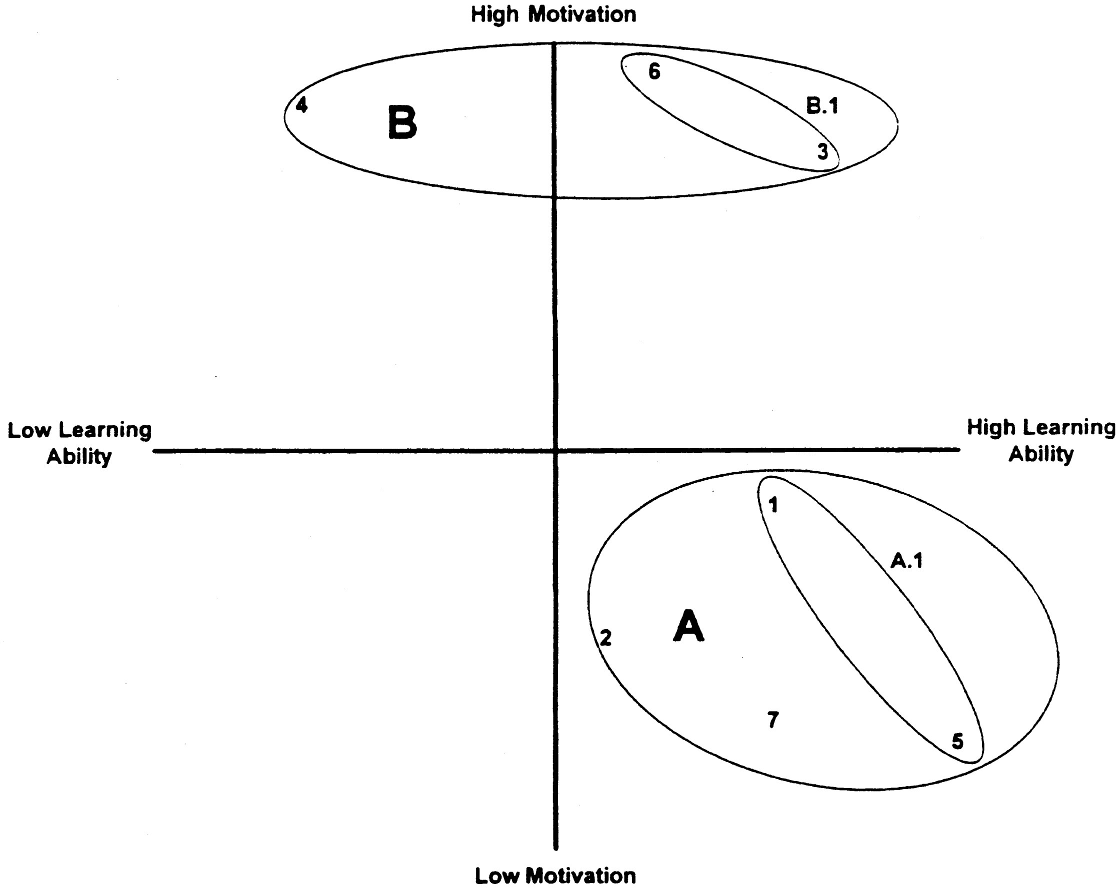

Figure 7. **Student clusters for Language Arts Tutor MR**

to a learning disability, most of MR's problem students were seen as capable but unmotivated. By contrast, MJ (like all math tutors in this study) saw *all* problem students, whether motivated or not, as lacking in capability. This difference probably does not reflect a difference in student samples, because MR and MJ tutored a number of the same students. It may reflect a difference in the way the tutors evaluate, but it possibly reflects a difference in how students are generally perceived in mathematics versus language arts. Poor math performance may often be viewed in our culture as both an ability and a motivational issue, whereas poor language arts performance may be viewed more often as a motivational problem primarily.

Discussion

Modeling Student Differences in Terms of Ability and Motivation

Using repertory grid interview techniques, five experienced tutors representing different domains and settings each identified the personal constructs they used to discriminate and categorize students for the purpose of adapting their tutoring to student differences. Cluster analyses of constructs for each tutor revealed two higher order dimensions of discrimination that seemed similar in meaning across all tutors. Based on study of the personal constructs that clustered together to define them, as well as corroborating evidence from interviews with tutors, we are confident that the main dimensions by which students are judged represent *motivation* and *ability*. Although other clusters appeared in tutors' analyses, the motivation and ability clusters were the most dominant for all tutors and they were the only ones that appeared for multiple tutors. Moreover, when tutors were asked to name their five most important constructs for tutoring, with only one exception, all tutors' top five constructs combined with either the motivation or the ability clusters. These findings indicate that, for the purpose of adapting tutoring to student differences, the most important ways that students differ from one another relate to their motivation and ability.

Individual tutors' concepts of *motivation* and *ability* were not exactly alike, but they shared numerous common features. For example, every tutor's motivation cluster contained one or more scales for rating students' enthusiasm for the learning task—scales based on such constructs as interest, curiosity, involvement, enjoyment, and engagement. Although they were distributed in different combinations across all five tutors, scales measuring reflectivity, frustration tolerance, persistence, attitude, and work habits were each found in the motivation clusters of two or more tutors. Many similarities across tutors were also seen in the ability clusters. Scales for assessing general ability or intelligence, learning speed, and mastery or achievement were each found in at least four different ability clusters, distributed across tutors in various combinations. Three ability clusters contained scales for judging student confidence, and two contained scales measuring students' degree of self-direction and degree of struggle. In sum, there was much evidence that motivation and ability were similarly interpreted and understood, as concepts, across tutors.

However, there were some differences in tutors' perceptions of these concepts, differences often influenced by tutors' personal views and biases, as well as by the types of settings in which they tutored. For example, like other tutors, KT's motivational cluster contained scales for judging students' activity and reflectivity, but it also included unique scales for judging the extent to which pairs helped one another and whether they followed instructions that were important to the experimental context of KT's tutoring. With respect to ability, KT was like other tutors in judging student competence as learning speed, degree of struggle, and mastery of basic skills. But because KT observed collaborative rather than individual learning, leadership ability was an important aspect of competence that was unique to KT.

In addition to judging students in terms of learning ability and motivation, tutors employed other personal constructs that were specific to their own style and influenced by the particular instructional situations and settings in which they worked. To the extent that such constructs are both tutor- and situation-specific, they are less universal. Moreover, tutors regarded these constructs as being less important, relative to ability and motivation, for adapting their tutoring to student needs. However, tutor-specific and situation-specific discriminations were useful for some tutoring decisions. For example, KT reported adjusting tutoring if students made common errors associated with particular problems.

In sum, there was considerable evidence that for all tutors, the most important differences among students for purposes of adapting tutoring to individual needs were differences in students' motivation and ability. These concepts were similarly understood, though not precisely alike, across tutors.

General Student Models as a Basis for Adaptive Tutoring

As a byproduct of judging and characterizing tutees on the basis of their motivation and ability during tutoring, all tutors grouped students into categories that can be described as general student models. These groupings were identified for each tutor by a cluster analysis of students. Comparisons of tutors with respect to how they grouped students revealed several interesting similarities and differences. Math tutors in both computer-assisted/experimental and noncomputer/natural tutoring environments possessed an implicit Problem Student group for tutees who were both unmotivated and less competent. Math tutors also had a large Other Student group that subsumed various subgroupings of students. For three math tutors, the Other Student group included categories that approximately represented motivated/competent, motivated/moderately competent, and motivated/less competent students. One math tutor (JS) also differentiated Typical Students, a group of moderately competent and moderately motivated tutees.

The structure of these clusters suggests that computer-based tutoring systems for mathematics should have a tutoring approach for dealing with Problem Students who are low in both motivation and ability. A separate tutoring approach for Other Students would be necessary. JS's world view implies that the dominant tutoring strategy for Other Students should be designed for typical (moderately motivated and moderately competent) tutees and should be adapted in response to increasing levels of competence and or motivation. A tutoring system based on the other math tutors' world view would have an Other Student tutoring strategy designed for motivated students and would be adapted in response to indicators of student competence.

Interestingly, no math tutor identified any student as both competent and unmotivated, raising the question of why this student category had no membership in the context of this study. Perhaps this category would have collected some competent-but-bored students if capable math students had not been challenged sufficiently by the tutoring tasks. Expert tutors probably avoid filling this category by conducting assessments and controlling the degree of challenge that students receive. The implication for computer-based tutoring is that this student model might be used as an assessment aid to help monitor for boredom. That is, if indicators place any student in the category *Competent-but-Bored*, a reasonable adaptive strategy could be to modify speed and challenge of tasks upward (thus, we hope, changing the applicable student model).

Several generalizations about student modeling and adaptive tutoring may hold true for tutoring in general, thus applying to the language arts, as well as the mathematics, domain. Most obvious of these is that similar dimensions of judgment (ability and motivation) appear to underlie modeling of student differences in both domains. However, MR differed from all math tutors in modeling most students as competent to write, with the major discriminating factor being their motivation to do so. From MR's perspective, the major tutoring challenge for most problem students was getting them to produce. For math tutors, however, there were two types of difficult instructional challenges: students who lacked both ability and motivation and students who wanted to perform well but were lacking in ability to do so.

In sum, our analyses provide evidence that tutors: (a) accumulate evidence about students during tutoring; (b) maintain cumulative models that characterize students generally in terms of their motivation and ability; and (c) adjust their tutoring to accommodate those general characterizations. Tutors reported adapting their tutoring to different students in the specific ways identified in Tables 4 and 5. However, longitudinal studies that include examination of tutorial dialogue over time are needed to verify that cumulative motivation and ability assessments are in fact salient to tutors during actual tutoring and to explain in greater detail how these are used in both local and global tutorial decision making.

The Implications of Tutoring Styles

The tutoring goals and adaptive modifications described by different tutors reflected different styles that varied in terms of content, philosophy, and complexity. This observation builds on the idea of Lepper et al. (1993) that tutoring styles constitute important forms of instructional variation that should be studied. Styles were not directly associated with student populations, content domains, or settings. For example, although they tutored different subjects and drastically different populations, the styles of tutors PM and MR were similarly indirect, therapeutic, and empathetic in nature, and both had the primary aim of developing student independence. For JS, the style was more laissez-faire for higher ability students, more didactic, direct, and structured for lower ability tutees. Thus, styles reflected between-tutor differences as well as within-tutor variations.

The concept of tutoring styles raises important questions related to interpreting, or re-interpreting, past research that has examined conversational structures and moves in tutoring. For, if there are within- and between-tutor style variations, the conversational moves and structures identified in past studies will be associated with some expert tutoring styles but not others. To illustrate, consider that research on remedial algebra tutoring by McArthur et al. (1990) reported that tutors spent significant conversational time in activities related to framing and managing tasks. In the current study, however, the amount and type of task framing and management were altered by some tutors as a function of student model. For example, JS completely controlled the selection of topics and activities for lower ability students, framing sessions for these students with review material. However, JS provided little task management for higher ability students, who were allowed to manage themselves, and lessons for the high ability were framed in terms of relatively advanced math-theory discussions. Though not previously mentioned, similar ability-related adaptations in task management also were reported by MR, who broke down tasks for the one student with a disability but merely reminded more capable students to manage their own tasks.

Other data from this study point to style-dependent differences in how responses to errors are handled. Our review of past tutoring research, reported earlier, led to the conclusion that error feedback is an important aspect of tutoring and that there are a variety of strategies for handling error correction. However, a difference between skilled and unskilled tutors was that skilled tutors often wait for students to correct errors, while unskilled tutors do not. In this study, error-response strategies were seen to be a function of tutoring style. For example, JS reported altering error-response strategy as a function of student ability in that higher ability students were expected to correct their own errors, while lower ability students were often corrected by the tutor. PM, whose tutoring style was different, also expected higher ability students to evaluate and correct their own mistakes, and he reinforced them for doing so. However, lower ability students, who did not spontaneously correct, were not corrected directly but were told that they relied on the tutor too much and were encouraged to evaluate their own errors.

These examples illustrate that certain conversational structures and moves change as a function of style, which is influenced by both pedagogical philosophy and perceived student differences. Future research on tutoring, including interpretive literature reviews, should examine such style differences more carefully.

The Importance of Motivation in Tutoring

In contrast to previous findings based on conversational analyses of adult human tutoring, the results of this study support the central tenet of Lepper et al. (1993) that student modeling by human tutors involves consideration of affective/motivational, as well as cognitive, states of students. All tutors watched for motivational indicators and adjusted their tutoring. Moreover, for three out of five tutors in this study, motivational/affective considerations were reported as more important than judgments of cognitive ability. Why use of motivational constructs has not been regularly observed in previous tutoring research is not altogether clear. Methodological bias may provide one explanation. That tutors do not explicitly reveal their personal constructs about students during their actual tutoring sessions with them is not surprising. Thus, conversational analyses of taped tutoring sessions that do not follow up with retrospective interviews of tutors would not reveal much personal construct psychology. Also, it is evident that personal constructs often influence between-lesson global decision processes, such as session planning, which might not be observed in studies that focus only on tutorial conversation. Finally, this study suggests that certain tutoring styles are generally more motivating than others. Tutors that know how to motivate may not see many motivational differences among students over time. This would be especially true in settings where tutees genuinely volunteer. In some natural settings, however, participation in tutoring is required if not coerced. As indicated in this study, motivation is a major factor in such environments.

It seems especially important to note that the role of motivation in student modeling has been observed only in studies that have relied substantially on interview data rather than pure conversational analysis. Although some conversational analyses have also employed retrospective interview in an effort to help uncover covert tutor thoughts, findings have sometimes been colored by the perspectives of researchers who asked directive interview questions. For example, if tutors are asked for the purpose of a decision, they may not be prompted to mention personal constructs. To illustrate, a tutor may assign a problem exercise to a tutee for the purpose of reviewing proportional reasoning and, if asked for the purpose, will state it as such. However, if the tutor were also asked, "Why did you select method A for student X instead of method B, which you used for student Y?" the tutor might respond that student X was less motivated and less competent than student Y, so method A would probably work better. This illustrates why it is important to examine tutor thinking using a number of different techniques.

Using the Evidence Tables

To the extent that the problem of intelligent tutoring is conceptualized as (a) observing evidence that can be used to evaluate students in terms of constructs that are relevant to tutoring and then (b) selecting tutoring responses, styles, or strategies that are appropriately adaptive to those evaluations, the findings of this research should be useful to developers of intelligent tutoring systems. Tables 4 and 5 were developed to summarize the types of evidence that three computer-assisted tutors collected to characterize students in terms of motivation and ability and what tutoring techniques or strategies were implemented in response to those evaluations. Asterisks in these tables indicate which items of evidence or tutoring strategies might be especially difficult for current technology to manage. As the low number of asterisks indicates, however, adaptive tutoring based on both motivational and ability indicators is feasible with current computer-based instructional technology. In fact, we have plans to implement some of these evidence-response relationships within the TiPS (Tutorials in Problem Solving) system, a computer-based tutor that has been used in adult learning environments.

Our system operates using a Bayesian inference network. Student behaviors believed to be reasonable indicators of the items of evidence listed in Tables 4 and 5 are observed by the TiPS system. In simple terms, evidence associated with high motivation or ability increases the probability that the student is a member of the model Exceptional (both highly motivated and competent). Evidence associated with low motivation and ability increases the probability that the student is a member of the group Problem Student (both unmotivated and not competent). Evidence associated with low ability and high motivation increases the probability the student is in the category Eager Student (motivated but not competent), and so forth. Different tutoring strategies and procedures associated with different student models can be triggered when probabilities reach threshold levels. The result is an adaptive assessment model in which the goal is to move students from less desirable to more desirable model spaces over time, based on accumulating evidence pertaining to motivation and ability in the specific tutoring context. We note, however, that this model has not yet been fully implemented or tested experimentally to determine whether adding adaptive tutoring based on motivation and ability assessments enhances the system's instructional effectiveness.

A Comment on Tutoring in Natural Versus Laboratory Environments

Comparisons of tutoring in natural, remedial environments versus laboratory, computer-based environments allowed us to estimate the aspects of adaptive tutoring that can and cannot be handled in a computer-based environment where there is little human support. When instructional computer systems are well designed and intelligent, little human intervention should be required in certain situations. For example, intelligent computer-based tutoring can be very suitable for adult workplace and continuing education programs where motivated students voluntarily participate during individually convenient hours for the purpose of self-improvement. When students are fundamentally self-monitoring and motivated to learn, they may not need the additional social support that was visible in the tutoring of MR and MJ. For students who lack motivation and related self-management skills, there are important components of empathic tutoring that can be described as parental in nature. These cannot be managed by a machine alone, even an empathic one.

APPENDIX

Interview Protocol for Tutors

Initial Knowledge Elicitation Phase

A. Brief Introduction to the Interview

1. Interviewer: "We are interested in how teachers/tutors think about their students. Do you have a group of students in mind which you can draw on for classification?"
2. Interviewer records the list of students and continues with the next

portion of the interview, the formation of a list of classification constructs.
3. Interviewer: "What was your main purpose, what were your objectives in the context of the tutoring you did with these students?"

B. Forming a List of Classification Constructs

1. The interviewer provides a randomly selected triad of students from the tutor's list. Interviewer: "Beginning with the first three students, name what you believe to be the most important attribute or trait that distinguishes any two members of this triad from the third one."
2. The interviewer asks the tutor to label the construct and its polar opposite. The tutor responds, for example, that, "Don and Chris are motivated; Terry is not." The tutor suggests that the construct should be motivated, along with its polar opposite, unmotivated.
3. Interviewer: "What specific evidence do you have to suggest that this student [the singled-out student] was (motivated, dull, cooperative, etc.) relative to others?"
   Interviewer: "What specific behaviors did this student exhibit that led you to believe him or her to be (motivated, dull, cooperative, etc.)?"
   Interviewer: "What specific behaviors did these two students exhibit that led you to believe them to be (opposite of just-mentioned construct)?"
4. Interviewer: "What other types of specific evidence would help you distinguish (motivated, dull, cooperative, etc.) from (opposite)?"
5. Interviewer: "If you had to tutor these three students in math problem solving, how would your strategy for this student [the singled-out student] differ from your strategy for the others?"
6. After all information has been recorded on the first triad of students, the interviewer again selects a random triad for presentation to the tutor. This cycle continues until several consecutive constructs given by the tutor repeat previously reported constructs and until at least 20 triads have been presented to the tutor. Data collected at this point should include the list of students for classification and a list of the classification constructs (including their polar opposites), all drawn from the expert tutor in his or her own language and terms.
7. The tutor is dismissed so that the interviewer can prepare the next part of the interview: constructing a rating grid from the list of constructs and the list of students.

Notes

Funding for this research was provided by the Office of Naval Research, Cognitive and Neural Science and Technology Division, Contract No. N00014-93-0310. The authors thank Sandra Marshall, Richard Lehrer, Bob Mislevy, and two anonymous reviewers for their comments on earlier drafts. We are also grateful to the unnamed tutors whose time and patience made this study possible. An abbreviated and earlier version of this analysis appeared in the 1996 Proceedings of the International Conference on the Learning Sciences, Chicago, IL.

References

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*(6), 4–16.

Boose, J. H. (1985). A knowledge acquisition program for expert systems based on personal construct psychology. *International Journal of Man-Machine Studies, 23*, 495–525.

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

Derry, S. J., Wortham, D., Webb, D., & Jiang, N. (1996, April). *Tutorials in problem solving (TiPS): Toward a schema-based approach to instructional tool design.* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Fox, B. A. (1993). *The human tutorial dialogue project.* Hillsdale, NJ: Erlbaum.

Goodyear, P. (1992). The provision of tutorial support for learning with computer-based simulations. In E. De Corte, M. C. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (pp. 391–410). New York: Springer-Verlag.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104–137.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1995). Hinting as a tactic in one-on-one tutoring. *The Journal of the Learning Sciences, 5*, 23–47.

Kelly, G. A. (1955). *The psychology of personal constructs.* New York: Norton.

Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*, 75–95.

Lepper, M. R., & Chabay, R. W. (1988). Socializing the intelligent tutor: Bringing empathy to computer tutors. In H. Mandl & A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 242–257). New York: Springer-Verlag.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 75–106). Hillsdale, NJ: Erlbaum.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction, 7*, 197–244.

Merrill, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction, 13*, 315–372.

Norusis, M. J. (1993). *SPSS for windows: professional statistics* (Release 6.0). Chicago, IL: SPSS.

Person, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A.C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction, 13*, 161–188.

Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal, 24*, 13–48.

Sleeman, D., & Brown, J. S. (Eds.). (1982). *Intelligent tutoring systems.* New York: Academic.

Van Lehn, K. (1988). Toward a theory of impasse-driven learning. In H. Mandl & A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 19–42). New York: Springer-Verlag.

Manuscript received April 23, 1996
Revision received January 21, 1997
Accepted February 10, 1997