COGNITION AND INSTRUCTION, *21*(3), 209–249

Why Do Only Some Events Cause Learning During Human Tutoring?

Kurt VanLehn, Stephanie Siler, and Charles Murray
Learning Research and Development Center
University of Pittsburgh

Takashi Yamauchi
Psychology Department
Texas A&M University

William B. Baggett
School of Computer Technology
Pittsburgh

Developers of intelligent tutoring systems would like to know what human tutors do and which activities are responsible for their success in tutoring. We address these questions by comparing episodes where tutoring does and does not cause learning. Approximately 125 hr of tutorial dialog between expert human tutors and physics students are analyzed to see what features of the dialog are associated with learning. Successful learning appears to require that the student reach an impasse. When students were not at an impasse, learning was uncommon regardless of the tutorial explanations employed. On the other hand, once students were at an impasse, tutorial explanations were sometimes associated with learning. Moreover, for different types of knowledge, different types of tutorial explanations were associated with learning different types of knowledge.

In principle, advances in Artificial Intelligence (AI) should make it easy to build tutoring systems that emulate human tutors. However, it is not yet clear what human tutors do. Although an initial picture exists due to existing studies, one goal of this research is to add more detail to this picture. We argue that it makes sense

Requests for reprints should be sent to Kurt VanLehn, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260. E-mail: vanlehn@cs.pitt.edu

to view tutoring as a sequence of *learning opportunities*, where a learning opportunity is a brief episode wherein the student and tutor apply one principle of physics to move the problem solving forward or discuss a principle, typically after the problem has been solved. If the student has already mastered this particular principle, then the student and tutor simply apply it. However, if the student is unfamiliar with the principle or only partially familiar with it, then the student and tutor struggle together with the dual objectives of getting the student to understand the principle and applying it to the problem. Although the learning opportunities for two different principles are often quite dissimilar, the learning opportunities for the same principle are often quite similar. For instance, tutors in our study tend to use the same explanations for a principle over and over.

A learning opportunity is only an *opportunity* to learn. Some learning opportunities seem to result in no learning. Thus, the second objective of this research is to characterize the differences between successful and unsuccessful learning opportunities. In particular, we test whether (a) learning is more probable when the student reaches an impasse, i.e., makes an error or gets stuck; (b) tutorial explanations have any impact; and (c) mentioning a problem-solving goal during a learning opportunity increases learning. These hypotheses are a first step toward characterizing which kinds of learning opportunities lead to success and which lead to failure. This information is critically important for the design of tutoring systems and other instruction.

Before explaining these research objectives in more detail, some background is required. In particular, the central concept of a learning opportunity must be carefully defined.

TUTORING: COACHED PROBLEM SOLVING

The dictionary defines tutoring as teaching with one student per teacher. Just as there are many different kinds of classroom teaching, so too are there many kinds of tutoring. The characteristics of tutoring depend on factors such as the material being taught, the student's prior knowledge and the tutor's pedagogical objectives, practices, and knowledge. This article is concerned with tutoring where:

- The students are learning how to solve mathematical, scientific, or technical problems that (a) require multiple observable actions, such as writing equations or making tests with a voltage meter, and (b) take many minutes to solve. Such learning is often called *cognitive skill acquisition* (VanLehn, 1996).
- The students have already learned some basic principles of the domain and have been introduced to the problem-solving process. In terms of Anderson's three stages of cognitive skill acquisition (Anderson, 1982; VanLehn, 1996), the students are in the second stage. They are neither total novices who are unable to

solve a single problem (stage 1) nor advanced students who are just improving their speed and reducing the probability of unintentional errors (stage 3). Stage 2 students still make errors caused by missing or incorrect knowledge, and they require help to solve some problems.

- The tutoring sessions consist of a sort of asymmetric collaboration where the tutor helps the student as the student solves the problem (as opposed to the student solving the problem, then showing the solution to the tutor for critiquing). This is often called *coached problem solving* (Shute & Psotka, 1996).

Perhaps because of the academic and economic importance of cognitive skills, many studies of coached problem solving have been conducted (Anderson, Farrell, & Saurers, 1985; Chi, 1996; Fox, 1993; Frederiksen, Roy, & Bedard, submitted; Heffernan, 2001; Hume, Michael, Rovick, & Evens, 1996; Lepper, Woolverton, Mumme, & Gurtner, 1993; McArthur, Stasz, & Zmuidzinas, 1990; Merrill, Reiser, Merrill, & Landes, 1995; Moore, 1993; Porayska-Pomsta, Mellish, & Pain, in press; Putnam, 1987; Schoenfeld, Gamoran, Kessel, & Leonard, 1992; Wood, Bruner, & Ross, 1976) and reviewed (Merrill, Reiser, Ranney, & Trafton, 1992; Shute & Psotka, 1996). Several commonalties have emerged.

First, most of the tutorial dialog focuses on steps in solving the problem, as opposed to general facts or principles in the problem domain, or metacognitive discussions of the student's knowledge. Moreover, the sequential and hierarchical structure of problem-solving steps determines the sequential and hierarchical structure of the dialog to a large degree (see especially Frederiksen et al., submitted). A short and hypothetical tutorial dialog for a classic cognitive skill, algebraic equation solving, adapted from dialogs analyzed by Neil Heffernan (2001) is provided here for illustration:

1. Tutor: Why don't you try problem 5 [points to $7 - 2x = 5x$]
2. Student: [writes $5x = 5x$]
3. Tutor: Umm…
4. Student: That looks funny.
5. Tutor: Yep. Better think again.
6. Student: [pauses 10 sec]
7. Tutor: Suppose you were evaluating $7 - 2x$ and x was 3 [writes $7 - 2x$ and $x = 3$]. How would you do it?
8. Student: 2 times 3 is 6, and 7 minus 6 is 1.
9. Tutor: Right. You did the multiplication first. You didn't do 7 minus 2 is 5, THEN multiply by 3 [points to the numbers as she mentions them]. So can you simplify $7 - 2x$ to $5x$?
10. Student: I guess not.
11. Tutor: Right. So how are you going to solve that equation?
12. Student: [Crosses out $5x = 5x$. Writes $7 - 2x = 5x$ then $7 = 5x - 2x$]

13. Tutor: That's better, but check your signs.
14. Student: Oops. [writes 7 = 5*x* + 2*x*]
15. Tutor: Good!
16. Student: [writes 7 = 7*x*]
17. Student: [writes *x* = 1]
18. Tutor: Excellent.

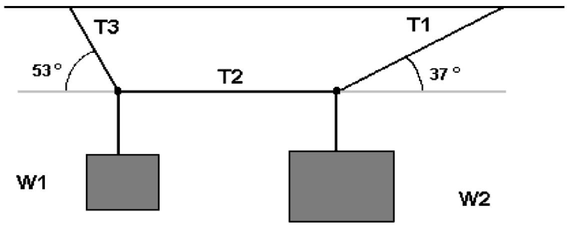
Much of this dialog is about the first step in the solution, which the student finally writes correctly at line 14. The student writes the other steps without difficulty. Although the discussion of operator precedence at lines 7 through 9 could be considered a digression because it is not directly addressing a step in the solution, it was clearly done as part of a tutorial teaching tactic (called the concrete articulation strategy by Heffernan). Most tutorial dialogues are exactly like this in that they stick closely to the steps in the solution.

A second common finding is that most tutors keep students on a correct solution path. If students make an error and do not detect it immediately themselves, the tutor almost always points the error out, sometimes just by pausing or saying "Umm…" as in line 3 (see especially Fox, 1993; Merrill et al., 1995). Although tutors seem to give immediate feedback naturally, some tutors deliberately delay feedback until a "quiet" period in the problem solving. For instance, Rovick and Michael, two often-studied expert cardiophysiology tutors, often deliberately delay their feedback, but even they will fall back on immediate feedback when students appear to be having trouble (Cho, Michael, Rovick, & Evens, 2000).

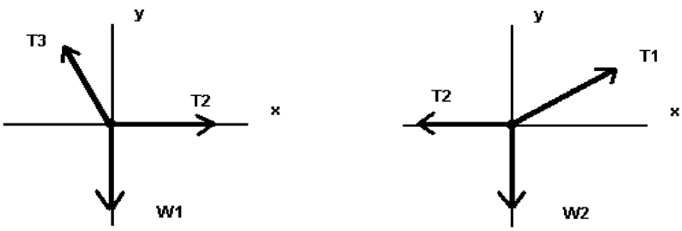
In summary, there is a consensus among investigators that, as Merrill et al. (1992, p. 300) put it, coached problem solving "can be viewed as a collaborative problem-solving effort, with each party contributing to the solutions," except that the tutor ensures that the solution does not stray very far from a correct solution path. Although a consensual picture of coached problem solving has emerged, it is not very detailed. More work is needed to understand tutoring in sufficient detail to constrain the design of intelligent tutoring systems. One goal of this research is to add detail to the emerging picture of coached problem solving. Thus, we studied college students being tutored with expert tutors.

Cognitive Task Analysis

The task domain studied here is a kind of physics problem solving that involves algebra and trigonometry but not calculus. Such problem solving is common in introductory college physics courses and advanced high school physics courses. A typical problem and its solution are shown at the top of Figure 1. The problem consists of a diagram and a statement, "The weight W1 in the picture is 300 N. Find T1, T2, T3, and W2." The solution process involves drawing vector diagrams



The weight W1 in the picture is 300 N.
Find T1, T2, T3 and W2.



For the Left Knot:

$\sum \mathbf{F} = 0$

$\sum F_y = T3_y + 0 - W1$

$T3_y = T3 \cos 37$

$W1 = 300$

$0 = T3 \cos 37 - 300\text{ N}$

$300\text{ N} / \cos 37 = T3$

$T3 = 375.6$

$\sum F_x = T3_x + T2 = 0$

$T3_x = -T3 \cos 53$

$0 = -T3 \cos 53 + T2$

$T2 = 226\text{ N}$

For the Right Knot:

$\sum \mathbf{F} = 0$

$\sum F_x = T2_x + T1_x$

$T1_x = T1 \cos 37$

$\sum F_x = -T2 + T1 \cos 37$

$T1 = T2 / \cos 37$

$T1 = 283\text{ N}$

$\sum F_y = T1 \sin 37 - W2 = 0$

$T1_y = T1 \sin 37$

$W2 = T1 \sin 37$

$W2 = 170.3\text{ N}$

FIGURE 1 A solved physics problem.

and writing equations (shown in the bottom of Figure 1). A competent student can solve such a problem in around 10 min, but the average student spends much longer. The multistep and multiminute nature of the solution are common features of all cognitive skills.

Physics problem solving has been studied often in both AI and cognitive psychology. AI problem solvers have been written to explore various cognitive task analyses (Bundy, Byrd, Luger, Mellish, & Palmer, 1979; de Kleer, 1975; McDermott & Larkin, 1978; Novak & Araya, 1980). Machine learning programs have shown how physics skill can be acquired from studying examples and solving problems (Elio & Scharf, 1990; VanLehn & Jones, 1993a, 1993b; VanLehn, Jones, & Chi, 1992). The differences between expert and novice physics problem solvers have been studied (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980; Priest & Lindsay, 1992) and modeled (Elio & Scharf, 1990; Larkin, 1981; Priest & Lindsay, 1992). It was in this task domain that the self-explanation effect was first uncovered (Chi, Bassok, Lewis, Reimann, & Glaser, 1989) and modeled (Reimann & Schult, 1993; Reimann, Wichmann, & Schult, 1993; VanLehn et al., 1992).

Although each of these works uses a different cognitive task analysis, they all seem to divide the knowledge required for physics problem solving into two types:

- Operational versions of the basic physics principles (e.g., Newton's laws).
- Knowledge about how to apply those principles to solve specific problems.

There is substantial agreement among the researchers not only on this division, but also on what the operational versions of the basic principles are. For instance, the physical law $W = m \cdot g$ appears prominently in all textbooks of physics. An operational version of it is: "If B is a body that is close to planet P, then $W = m \cdot g$, where W is the magnitude of the force on B due to gravity, m is mass of B and g is the gravitational constant of planet P." What makes this version of it operational is that it includes the conditions for the applicability of the law and the definition of its symbols. All the fundamental laws in the textbooks are easily converted to operational versions.

Some problem solutions require operational versions of "principles" that do not appear in the textbook. For instance, one such tacit principle is, "If two objects move together, then they can be treated as a single body." Although physicists would certainly not accord the same status to such tacit principles as they do to Newton's law and other principles that do appear in textbooks, the model-builders mentioned earlier accord all the operational principles essentially the same role in problem solving and they use the same formal representation for all operational principles.

Computational studies indicate that some kind of knowledge beyond the operational principles is used by human solvers. If an AI problem solver has only the operational principles, then it must search extensively to find a solution even if its knowledge of the operational principles is flawless and complete. For instance, if the solver uses the general-purpose method called *working forwards*, wherein it starts with the given information in the problem and augments it by applying operational principles, then it will often produce equations that are useless for deriving the particular quantity that the problem seeks. On the other hand, if the solver applies the *working backwards* method, wherein it starts from the sought quantity and works backwards toward the givens, then it will sometimes travel down dead end paths, producing useless equations before connecting to the givens. In contrast to AI problem solvers, students who have learned how to solve physics problems seldom produce useless equations (Priest & Lindsay, 1992). Clearly, they learned more than just the operational principles that are applied by general-purpose methods such as forward or backward search. This extra knowledge tells them which principles to apply and which to ignore.

Although researchers agree that some kind of extra knowledge beyond the operational principles is required, there is little consensus on how to represent it. Some researchers favor case-based representations (Elio & Scharf, 1990; Reimann & Schult, 1993; Reimann et al., 1993). Others favor schema-based representations (Bundy et al., 1979; Chi et al., 1981; Larkin, 1983, 1981). Others prefer finer-grained search control knowledge (VanLehn & Jones, 1993a; VanLehn et al., 1992). All these studies involve building computational models and comparing them to protocols of human problem solving. Although one can sometimes discriminate among proposed knowledge representations by considering how they could be learned, all these researchers have also built models of how their extra knowledge is learned from problem-solving experience. Thus, there is as yet no consensus on how the extra knowledge is represented.

However, one small point that all researchers agree upon is that this extra knowledge is responsible for managing problem-solving goals, where a goal is the student's intention or commitment to do certain actions that lead to a certain state (Pollack, 1990). For instance, typical physics problem-solving goals are to draw vectors for all the forces acting on a certain object, or to determine the value of a certain variable. Because there is no consensus on what this extra knowledge is except that it involves goals, we call it *goal-management knowledge* to distinguish it from the operational versions of the principles.

In short, there is agreement that the knowledge required to solve physics problems consists of operational principles and goal-management knowledge. There is agreement on what the operational principles are, but there is no agreement on what the goal-management knowledge is.

However, these authors do agree that the knowledge required for solving physics problems is only a small fraction of the total knowledge that physics students should learn. For instance, consider the operational principle, "If there is a force F1 acting on body X due to body Y, then there is also a force F2 acting on body Y due to body X. The forces are called a *reaction pair*, and they have equal magnitudes and opposite directions." This is an operational version of Newton's third law, which is usually stated, "For every force, there is an equal and opposite reaction force." To be fully competent, students must understand that the operational version is in fact a version of Newton's third law, that the law was first formulated by Issac Newton, and that Newton's laws are empirical principles that are supported by countless experiments. In fact, deeper reflection on the third law might entail realizing that one force does not cause the other force, but instead they are the same thing viewed from two different perspectives. In short, fully understanding the concept of reaction forces requires more than mastering this one operational principle. This example illustrates that the operational principles are only a fraction of the total knowledge of physics.

It also illustrates the dual relationship of concepts and principles. That is, concepts like *reaction force pair* are mentioned by operational principles and other pieces of knowledge, so mastering a concept means understanding all the principles, empirical findings, history, and other knowledge that mention that concept.

A major advantage of choosing physics as a task domain is that there is at least some agreement on how to analyze the knowledge required for solving problems, namely, that it consists of operational versions of principles and goal-management knowledge. However, it should be mentioned again that these two kinds of knowledge are only a part of the overall knowledge required of competent physics students.

Objectives of the Study

To focus this research, we examine how students learn operational versions of principles and ignore their acquisition of goal management knowledge. Because we seldom have occasion to talk about principles in other than their operational form, we use principle to mean the operational version of the principle.

As a second simplification, we assume that each principle is learned independently of the other principles and of the goal management knowledge. For instance, if the tutor and the student are not talking about Newton's second law or using it even implicitly, then it is unlikely that the student's understanding of Newton's second law will be influenced by the conversation. The independence assumption is probably only partially true. We adopt it to simplify the analysis. If principles are learned independently and we are trying to account for how the student learned a certain principle, Newton's third law, for example, we need only look at the episodes in the tutoring session where that principle is being discussed or applied.

An episode where a principle is being discussed or applied is called a *learning opportunity* for that principle. It is easy to identify the learning opportunities for a principle. A coder who knows physics can see explicit discussions of the principle in the transcripts, and can infer applications of the principle during periods of silence by seeing what equations and vectors the student writes. Although the tutoring certainly presents opportunities for all kinds of learning, we use the term *learning opportunity* only for episodes that could increase the student's knowledge of a principle.

To summarize, we view the tutorial dialog as a sequence of learning opportunities interspersed with discussion of other topics. Each learning opportunity addresses one principle, and there may be more than one learning opportunity per principle. A fairly complete picture of how a principle is learned can be obtained by analyzing the learning opportunities for that principle.

Given these simplifications, the objectives of the study can be stated more precisely. The first objective is to add detail to our picture of coached problem solving. Although we already know something about its overall structure (e.g., that the dialogue follows a correct solution path, and has the same sequential and hierarchical structure as that path), we do not know much about what occurs during learning opportunities. Just as different problems have different solution paths, we expect that the learning opportunities for different principles might have different structures. Thus, we collected learning opportunities for each of five principles, and determined five structures, one for each principle.

The second objective of the study is to determine what features of learning opportunities were associated with learning. That is, for each of several principles and several students, we first determined whether that student acquired the principle during the tutoring or not, then we collected all the learning opportunities the student had for that principle and coded those episodes of the dialogue. Using three different ways of aggregating these data, we looked for statistically reliable associations between the codes and the outcome: whether the relevant principle was learned or not.

METHOD

Participants

Two experienced adult physics tutors were recruited. Both had advanced degrees in physics and had served as assistants in multiple physics courses. The tutees were 42 students recruited from the University of Pittsburgh's calculus-based introductory physics course. All had been taught kinematics and dynamics earlier in the semester, so these tutoring sessions were partly a review for them. Students were compensated for their participation with extra course credit and money.

Materials

Five physics problems were developed from problems used in earlier studies and pilot tested on advanced high-school students. The problems involved only straight-line kinematics and dynamics, which are the first two topics taught in introductory physics courses. Figure 1 shows one of the problems used.

The five physics problems were analyzed to determine which principles were required to solve them correctly. The physics principles are similar to the ones used in Cascade (VanLehn et al., 1992), a computational model of cognitive skill acquisition. Cascade was successfully evaluated by comparing its reasoning to thousands of lines of protocol data. Not only did the participants' reasoning correspond to the reasoning generated by the principles (VanLehn & Jones, 1993b), but their learning events also corresponded to learning of individual principles (VanLehn, 1999). Because the principles used in this study are similar to the ones used in Cascade, it is plausible that these are indeed the principles that students learned and applied. As mentioned earlier, all models of physics cognition have included principles in their knowledge representation, and the particular principles they included are similar to the ones we used. This is not surprising, given that most of the principles are presented in physics textbooks, albeit often in nonoperational form.

Based on the analysis of the training problems, a test was developed whose items were designed to determine if a student had learned a particular principle. Each test item presented a situation where the principle should apply, and posed a goal that matched the principle. For instance, for the operational principle "If there is a taut string or string-like object attached to a body, then it exerts a tension force on the body," one test item could be:

> Suppose a satellite is being reeled into the space shuttle by a thin, nearly massless string. Suppose we idealize the situation and treat it as just three interacting particles: the satellite, the string and the space shuttle. Please draw all the forces acting on these bodies, and explain why each exists.

If the student has mastered the principle, then she should draw a tension force acting on the satellite and another tension force acting on the space shuttle. This same test item can also be used for testing knowledge of reaction forces. Such test items are like steps that would occur during normal problem solving. However, the item is written to articulate clearly both the inferences that would normally have occurred before this step (e.g., that the satellite, string and shuttle should be idealized as particles) and a goal for the step (e.g., to draw all the forces on all the bodies). Thus, all the cues required for retrieval and selection of the principle are explicitly present in the problem statement. This means that the student's knowledge of the principle can be revealed even if the student knows no other principle. On a more conventional test, ignorance of one principle might suppress the cues needed for retrieval and selection of a different principle. Thus, our test was designed to measure knowledge of each principle independently.

Procedure

The students took a pretest, solved all five problems with the tutor's help, and then took an identical posttest. The tests took about a half-hour each, and the tutoring session took about 3 or 4 hr, with periodic breaks. The tutoring sessions were audiotaped and later transcribed.

To increase the applicability of this study to computer-based tutoring, the students solved physics problems with a computer program that allowed entering standard physics notations: vector drawings, algebraic equations, and text. The program was not a computer-based tutor; it served only as a recording and display device with domain-specific drawing tools.

To facilitate transcription, the student and tutor were placed in different rooms. They communicated by phone; the tutor watched a copy of the student's screen but could not manipulate it. This prevented them from communicating nonverbally. In face-to-face tutoring, Fox (1993) and others have found that gaze direction, gestures, and facial expressions communicate significant information, so transcription and analysis of face-to-face conversations must include much more detail than transcription and analysis of audio-only conversations. At any rate, neither students nor tutors found it difficult to use the linked screens and the phone, so the sacrifice in verisimilitude seemed to have been worth the resulting facilitation of transcription and the similarity to computer-based tutoring.

Analysis of Test Data

For each student and principle, both pretest and posttest were scored according to whether the student used the principle. Because the test problems were designed to be solvable using only one or a few principles each, interrater agreement was high (2 coders, $\kappa = .95$).

If a participant failed to use a principle on the pretest but used it on the posttest, the participant was said to *gain* that principle. If a participant failed to use a principle on both the pre- and posttests, then the participant was said to *not gain* that principle.

THEORETICAL ISSUES THAT MOTIVATE THE CODING CATEGORIES

Because our objective is to find features of learning opportunities associated with gains, we chose coding categories (features) for which there is some reason to expect an association with gain. This section motivates the particular features we chose to examine.

Impasses

Earlier work on cognitive skill acquisition without the aid of a tutor found that learning was often associated with impasses (Brown & VanLehn, 1980; Carroll & Kay, 1988; Siegler & Jenkins, 1989; VanLehn, 1987, 1990, 1999; VanLehn & Jones, 1993b; VanLehn et al., 1992). A nonoperational definition of an impasse is that it occurs when a student realizes that he or she lacks a complete understanding of a specific piece of knowledge. However, because impasses are usually coded from verbal protocols, a more operational definition is that an impasse occurs when a student gets stuck, detects an error, or does an action correctly but expresses uncertainty about it. The basic idea is that an impasse motivates a student to take an active role in constructing a better understanding of the principle. The student may interrogate memory, objects in the local environment (e.g., a textbook), or nearby people.

In contrast, a nonimpasse occurs under three conditions: (a) the student applies a piece of knowledge correctly with no signs of uncertainty, (b) the student makes an error that is never detected, or (c) the student somehow avoids applying the piece of knowledge, for example, by getting someone else to apply it.

In the tutoring context, almost all errors are detected, so that leaves the following coding categories:

- The tutor applies the principle before even giving the student a chance.
- The student tries to apply the principle and reaches an impasse by either
  —applying the principle incorrectly,
  —getting stuck and either asking for help or pausing for a long time, or
  —applying the principle correctly while expressing doubt or asking for confirmation.
- The student applies the principle correctly with no questions, pauses, or other signs of uncertainty or confusion.

We hypothesize that a student's understanding of the principle usually increases if the student reaches an impasse. In particular, if the tutor applies the principle, then the student's state of understanding is usually unchanged, even though the tutor might accompany the demonstration with considerable explanation. On the other hand, if the student applies the principle correctly, then the student apparently already understands it sufficiently well, so there is usually no further increase in understanding even though such practice tends to cause increases in speed and accuracy (Newell & Rosenbloom, 1981). Because there are already well documented cases where understanding increased with no sign of impasses (Jones & VanLehn, 1994; Siegler & Jenkins, 1989; VanLehn, 1991), we expect this hypothesis to emerge as a probabilistic association rather than an inviolate law of learning.

Explanations

We use explanation for both discussion of the principle itself, either in its generic form or its operational form, and discussion of how to apply the principle to the current problem state. An explanation can be produced collaboratively, with contributions by both the student and the tutor. In particular, tutors often try to elicit knowledge of the principle by mentioning a small part of it. For instance, if the principle is Newton's third law, then a hint such as "Are there any reaction force pairs here?" would count as an explanation because it mentions a part of the principle. (The tutor probably hopes that the student will fill in the rest.) On the other hand, merely giving the student feedback (e.g., "Not quite. Try again.") would not count as explanation, because it mentions no part of the principle or its application. Eliciting a goal (e.g., "Are you trying to find forces on the block?") usually would not count as explanation, but it could if the elicitation contained some information about the principle (e.g., "I'm assuming you are applying Newton's law, so are you trying to find forces on the block?"). In short, the definition of explanation is based on the *content* of a discourse segment rather than the identity of the speakers uttering it or the segment's function in the problem solving.

In addition to mentioning the principle or its parts, tutors sometime explain principles by deriving them, either mathematically or qualitatively. For instance, suppose the principle is, "if velocity is constant, then acceleration is zero." The tutor might start out by asking, "What is the definition of acceleration?" and hoping that the student will respond correctly with "Change in velocity divided by duration." The tutor may then assert the second step of the derivation and elicit the third step by asking, "So if there is no change in velocity, as is the case here, what is the acceleration?" On the other hand, simply encouraging the student to derive the principle would not count as an explanation, because it mentions no part of the principle or its derivation. Chi, Siler, Jeong, Yamauchi, & Hausmann (2001) called these "content-free prompts."

A major issue in the literature is whether learning is best accomplished with no explanation, student-generated explanation, or tutor-generated explanation. When computer-based tutors told the student that they had made an error, giving some explanation was better than giving no explanation, but the benefit was surprisingly small (Anderson, Conrad, & Corbett, 1989; McKendree, 1990). In studies with human tutors, it appears that student-generated explanations (self-explanations) are more effective than tutor-generated explanations (instructional explanations) (e.g., Brown & Kane, 1988; Chi, 1996), Moreover, having students self-explain but ask for instructional explanations when they reach impasses seems even better than self-explanation alone (Renkl, 2001). This would suggest the following ranking: No explanation causes the least learning, followed by tutor-generated explanations, then self-explanations, and finally by self-explanation with tutorial explanation as backup. However, when Chi et al. (2001) directly compared tutoring that contained mostly tutor-generated explanation with tutoring that contained mostly student-generated explanation, they found that students learned equally well in both conditions. Given this recent interest in tutorial explanations, many of our codes quantify how much explanation was given and by whom.

Goal Setting

When an intelligent tutoring system gives a sequence of hints to a student, the first hint usually states a goal, such as "Try drawing the acceleration of the car," that the tutor thinks is appropriate at this moment (Corbett, McLaughlin, & Scarpinatto, 2000; Shute & Psotka, 1996). Such goal-setting hints should be useful both because they insure that the student and the tutor are working on the same goal, and because the goal is a retrieval cue and a selection cue for the principle that the student should apply, so getting it into the student's focus of attention during a learning opportunity ought to encourage the formation of useful associations. Not surprisingly, such goal-setting hints sometimes work, and no further hints are needed in order for the student to do a correct action (Aleven & Koedinger, 2000). Even if they do not work, they set the stage for discussion of the principle, and thus are likely to increase the chance that the student will learn the principle. As this logic would predict, McKendree (1990) found that hints that mentioned goals are marginally more effective than hints that do not. This suggests that human tutors should also mention goals, either by telling them to the students as many tutoring systems do or, more likely, by eliciting them from students by, for instance, just asking them, "what are you trying to do here?"

One existing analysis of human tutoring examined whether human tutors mentioned or elicited goals. McArthur et al. (1990) found that only one of their three tutors mentioned goals. Moreover, that tutor's discussion of goals varied in a predictable way. Initially, the tutor mentioned goals frequently ("modeling" the goal management, according to McArthur et al.) but the frequency of goal mentioning declined rapidly ("fading," according to McArthur et al.). This pattern led them to speculate that the other two tutors did not mention goals because they felt their students had already mastered the relevant goal management knowledge, and thus did not need to mention goals.

At any rate, we suspect that mentioning or eliciting goals could help learning in some cases, so we coded for them in analyzing the transcripts. Thus, the three major categories of codes were:

- Was the first application of the principle done by the tutor, an impasse by the student, or no problem for the student?
- Was the principle explained in some way by either the student or the tutor?
- Was an appropriate goal mentioned by either the student or the tutor?

In addition to these three major categories, we included a few codes (e.g., the number of words in the learning opportunity) that did not fit into any of these categories.

Preview of the Three Analyses

We made three attempts to determine what features of tutorial dialog were associated with learning. All three used coding categories motivated by the theoretical issues described earlier. That is, we coded learning opportunities for impasses, explanations, and goals. In the first analysis, regression produced some interesting patterns, but did not reveal the impact of impasses, explanations and goals on learning. The problem seemed to be that we were treating all learning opportunities the same even when they addressed different principles.

The second attempt avoided this problem by selecting five principles and analyzing their learning opportunities separately. This allowed us to increase the specificity of the codes, as well as to add some new codes to address a few other issues that have been raised in the literature, namely:

- If the student made an error, did the explanation focus on the error, on the action that should have been done, or both? Sleeman, Kelly, Martinak, Ward, and Moore (1989) showed that tutoring on the correct action is just as good as first explaining why an error is wrong then tutoring on the correct action.
- At what level of generality was the discussion? Studies of analogical transfer indicate that having the instruction describe the principle in general terms, as opposed to letting students form those generalizations themselves, facilitated students' application of the principle later (Catrambone, 1994a).
- If the principle is mathematical, does the explanation consist of a derivation of it or just restating it in words? Studies of analogical transfer indicate that instructions that contain derivations facilitate transfer (Catrambone, 1994b).

Although this second attempt revealed several interesting patterns, it still did not address our main questions about the effects of impasses, explanation, and goals. The problem was that the learning opportunities for different principles were so different from each other that sometimes there would be zero occurrences of an event for one principle, when that event was common with another. These zeros prevented certain trends from emerging in the data analysis.

Thus, our third attempt involved formulating these trends as hypotheses and testing them using data from all five principles at once. This approach succeeded in revealing how impasses affect learning in this context, although the effects of explanations and goals still remain unresolved.

FIRST ANALYSIS: WHAT GENERAL FEATURES OF LEARNING OPPORTUNITIES CORRELATE WITH GAIN?

Our first analysis sought features that correlated with gain regardless of which principle was involved. Because we could not feasibly analyze all 42 students, we chose a subset of 8 that would maximize the contrast between effective and ineffective learners. For each of the two tutors, we selected for analysis two students with high gains and two students with low gains, balancing for pretest score and gender.

To collect learning opportunities, we first located each principle that was missed on the pretest. Different students missed different principles, of course. A total of 66 principles were missed on the pretest (i.e., each student missed about eight principles), so there were 66 student—principle pairs for analysis. Of these 66 pairs, 23 were gain (i.e., the student got the principle right on the posttest) and 43 were no gain. Thus, the challenge was to find features of the learning opportunities that were present for the 23 gain and absent for the 43 no gain.

For each of the 66 participant—principle pairs, all learning opportunities were located in the protocols. An episode was classified as a learning opportunity if the tutor and student discussed the principle or one of them applied the principle.

The learning opportunities were analyzed using the codes shown in Table 1. As discussed earlier, we believe insights into learning can result from analyzing learning opportunities in terms of impasses, explanations and goals.

TABLE 1
General Codes

Initiation Codes

- Did the student get stuck or make an error?
- Who initiated the discussion: Student or tutor?
- What initiated the discussion: The student detecting an error, the tutor detecting an error, or the student getting stuck?

Explanation Codes

- How many of the key ideas behind the principle were mentioned by the tutor? By the student?
- How many misconceptions were mentioned by the student? By the tutor?
- Who first mentioned the correct conclusion generated by applying the principle: The student, the tutor, the tutor by splicing words into the student's incomplete conclusion (Graesser, Person, & Magliano, 1995), the student when there was only one plausible choice, or neither?
- How many times was the correct conclusion mentioned by the student? By the tutor?

Miscellaneous Codes

- Was the action performed by the student without help from the tutor, by the tutor without help from the student, or by both?
- How many words were uttered by participants during the discussion?

However, there were few explicit comments about goals, so we did not bother to code for them in this analysis. Thus, Table 1 is divided into sections for codes concerning what initiated the event (e.g., an impasse), codes concerning the content of the explanations and who articulated that content, and miscellaneous codes.

Objective codes (e.g., number of words) were scored by one coder. Subjective codes were judged by two coders. Most codes were moderately reliable ($\kappa > .68$ for categorical codes; $r > .69$ for numerical codes), and those that were not reliable were discussed and re-coded.

Because one of our research questions is "what features distinguish effective learning opportunities from ineffective ones," the unit of analysis should be learning opportunities. However, if a particular principle is discussed several times by the student and the tutor, then there are several learning opportunities for it. However, we only have tests at the beginning and the end, and not in between each learning opportunity. Thus, we can only determine whether a set of learning opportunities was associated with gain or no-gain of that principle for that student. Thus, it was necessary to aggregate the results of the coding, which we did as follows:

- If the code was categorical (e.g., Did the student initiate the event?), we counted the number of learning opportunities with that code.
- If the code was numerical (e.g., How many misconceptions were mentioned by the student?), we summed over the learning opportunities.

As a result of this aggregation, we obtained 66 data points, each bearing value for a single binary dependent variable (gain vs. no-gain) and multiple integer independent variables (the codes listed earlier).

Because some data points come from the same student, albeit with different principles, it could be argued that they are dependent, which would complicate the statistics. However, we are assuming that principles are acquired independently, as discussed earlier.

Regression Results

Because the dependent variable was dichotomous, we used logistic regression. When regressed individually, only three codes were significantly correlated with gain:

- Did the student perform the action without help from the tutor ($r^2 = .030$, $p = .034$)?
- Did the student first mention the correct conclusion generated by applying the principle ($r^2 = .034$, $p = .026$)?

- How many words were uttered by student and tutor together ($r^2 = .075$, $p = .004$)? Events with fewer words more often resulted in gain.

The first two codes suggest that students who do more reasoning themselves learn more, as the self-explanation effect suggests (Chi et al., 1989). However, principle complexity could also explain the first two correlations. That is, simple, easily learned principles might be more often performed by the student without help and might tend to have their conclusions mentioned first by the student. This suggested analyzing the data with the intrinsic difficulty of learning the principles factored out.

Regression Results Factoring Out Intrinsic Difficulty

Because we do not know how to measure the intrinsic difficulty of learning a principle, we used two different approximations: the frequency of gain of the principle and the number of key ideas underlying the principle. The frequency of gain of a principle is the proportion of students who gain the principle among those who had an opportunity to gain the principle. That is, it is the number of students who got the principle wrong on the pretest and right on the posttest divided by the number of students who got the principle wrong on the pretest.

Although frequency of gain is certainly sensitive to the intrinsic learnability of the principle, it is also sensitive to the pedagogical techniques that tutors have for that principle. For instance, if two principles have the same intrinsic learnability but tutors have better explanations for one than the other, then the principles will have different frequencies of gain. Thus, we also used the number of key ideas underlying a principle, which is a measure that depends only on the content of the principle. The key ideas of a principle include a precise statement of the principle, definitions of any technical terms used in the principle and other propositions that are strongly related to the principle. For instance, the following key ideas underlie the principle "If an object is moving in a straight line and slowing down, its acceleration is opposite its velocity." The first key idea is a precise statement of the principle, and the other two are definitions of concepts used in it:

- If a body is moving in a straight line and its velocity is decreasing, the direction of its acceleration is opposite of the direction of its velocity.
- Average acceleration is change in velocity divided by change in time.
- The sign of a vector component specifies its direction relative to an axis.

Although the exact content of the key ideas did not matter for this analysis, the number of key ideas was used as a crude measure of the intrinsic learnability of the principle.

To factor out the learnability of principles, we first used the frequency of gain of principles as a measure of learnability. When gain was regressed against both frequency of gain and each of the learning opportunity codes, no code was reliably associated with gain. Interactions between the codes and frequency of gain were also not significantly associated with gain. Stepwise regressions showed that only frequency of gain was significantly correlated with gain and it explained 45% of the variance. In short, when difficulty is measured by the frequency of gain of principles, it is so strongly associated with gain that no other features of the tutorial dialogs reached significance after it was factored out. This is probably due to the fact that frequency of gain measured not only the intrinsic difficulty of the principle but principle-specific tutorial tactics. If it includes almost everything that could be expected to influence gain and we factor it out, then there is no variance left to explain with the codes.

The second measure of intrinsic difficulty, number of key ideas, is based solely on the content of the principle and thus is probably a better measure of learnability, although clearly not perfect. When gain was regressed against both the number of key ideas and each of the learning opportunity codes, only one code remained significantly associated with gain, namely, the number of words in the learning opportunity ($r^2 = .045$, $p = .015$). We examined interactions between pairs of categories that were significantly correlated with the dependent measure in the earlier analysis, and found that none of the interactions were significantly associated with gain. We performed a stepwise logistic regression, entering significantly correlated variables from the earlier analysis and their interactions, ordered by the p-values of the univariate tests from the earlier analysis. The selected model included the number of words ($r^2 = .043$) followed by an interaction term (number of key ideas X student first mentioned the correct conclusion; $r^2 = .011$). Thus, when we factored out the “intrinsic difficulty” of the principle, the only reliable correlate of gain was the number of words uttered during the learning opportunities. The fewer the words, the more likely the students were to gain.

Discussion

The association between verbosity and gain could be due to several factors. It could be that our use of the number of key ideas to measure intrinsic learnability was not entirely accurate, and that difficult-to-learn principles were associated with both long discussions and lack of gain. It could be that some event either before the discussion or early in it confused the student, which led to both a longer discussion and a lack of gain. It could also be that when tutors offered wordy explanations, students failed to learn, as Lewis (1989) found (cited in Anderson, Corbett, Koedinger, & Pelletier, 1995, p.191).

Although an association with verbosity was found, other possible correlates of learning may have been washed out by variations in learning opportunities caused by the principles themselves. Perhaps the activity during a learning opportunity depends more on the nature of the principle than on the learning strategies of the participants or the tutoring strategies of the tutors.

SECOND ANALYSIS: FOR EACH PRINCIPLE, WHAT FEATURES WERE ASSOCIATED WITH GAIN?

We suspected that our first analysis did not reveal strong explanatory patterns because it sought a single pattern that explained gain for every principle. Our next analysis thus examined principles separately, looking for features of the tutorial dialogs that explained just the gain on that principle.

Because we needed enough learning opportunities per principle to get statistical power, we abandoned the laborsaving device of examining only 8 participants and analyzed all 42 participants. However, we did not examine all principles. We examined a principle only if 5 or more participants gained on it and 5 or more participants failed to gain on it. These constraints were necessary to have enough variance to explain. The constraints eliminated all but five principles.

It seems plausible that there are four general influences on whether the student learns a particular principle:

1. *Method*: What occurred during the learning opportunities? This influence is the one that we are most interested in. Its analysis is presented later.
2. *Content*: What principle is being learned, and in particular, what is its intrinsic learnability? By analyzing each principle separately, we control for these influences.
3. *Teacher*: Who was the tutor, and in particular, what is the overall competence of the tutor? For each principle, we separately checked whether the identity of the tutors could explain the gains. For none of the 5 principles were gains associated with tutor (chi-square test, N = 15, 23, 32, 25, and 18).
4. *Student*: Who was the student, and in particular, what is the overall competence of the student? We also checked whether gains on a principle could be explained by the overall competence of the students. For instance, for the deceleration principle discussed later, of the 15 students who missed the principle on the pretest, 8 gained and 7 did not, and the pretest scores of the gainers (25.4) and nongainers (27.1) were not significantly different (p = .57). In fact, for none of the 5 principles were the pretest scores of the gainers significantly different from pretest scores of the nongainers (t test, N = 15, 23, 32, 25, and 18).

In short, it appears that learning is associated with what occurred during the tutorial dialog, and not with who was involved. Each of the following sections discusses one principle and the tutorial dialog features associated with its gains. All statistical tests were done with Fisher's exact test unless otherwise noted. For each principle, we begin by describing the learning opportunities themselves in some detail to put the rest of the analysis in a meaningful context.

As discussed earlier, the codes were designed to describe the goal setting, initial application of the principle, and its explanation. If a particular principle had no variance for a code (e.g., all the learning opportunities were initiated exactly the same way), then we usually did not include that code in the tables, but instead mention the uniformity in the text.

The Knot Principle

Students often think that the only objects to which they should apply Newton's law are blocks and other objects with mass. However, for some problems, massless objects are the appropriate choice for the body. (Physicists use "body" to mean the object to which one will apply laws.) A common massless object is a knot formed by tying together several massless strings. Ideally, the principle to be learned is, "A massless object can be a body." However, the only massless objects used in our problems are knots, so students may have learned only the more specific principle, "A knot can be a body."

In the training, this principle was used on a problem where two blocks are hung from a harness of five massless strings that has two knots (see Figure 1). The correct solution follows from applying Newton's law once for each knot. In the testing, the knot principle was used on a problem where two men are pulling a cart with a harness that has three strings and one knot. Thus, students must transfer the application of the principle from a vertical case to a horizontal case, and from a more complex harness to a simpler one.

Description of the knot principle's learning opportunities. Tutoring on the knot principle proceeded as follows. Tutors sometimes mentioned quite early in the problem that there were knots at the junctions of the strings, but they did not at that time mention that knots could be bodies. When they came to the part of the problem where a body needed to be chosen, they either explicitly stated the goal that a body must be chosen (20 cases) or did not explicitly state this (12 cases). If the tutor stated that a body must be chosen, then the tutors usually also asked the student to choose the body (16 cases). Regardless of how the body choice was prompted, either the tutor chose the body (11 cases), the student chose the wrong body (14 cases), the student chose the right body (5 cases) or the student could not choose a body (2 cases). When the tutor chose the body, it was often indicated implicitly by simply asking the student to draw forces on the knot. When the student chose the wrong body, the tutor always pointed it out immediately. Regardless of who chose the knot as the body and how, in 13 cases the tutors explained the principle. During six cases of those explanations, they stated a general version of the principle, such as "A body should be chosen that connects objects with known properties to objects whose properties we seek."

Features associated with learning the knot principle. Table 2 shows the codes that were used. The following features were associated with gain:

- Whether a student incorrectly chose the body (14 cases) or not (18 cases, $p = .013$).
- Whether the tutor asked the student to choose a body (16 cases) or not (16 cases, $p = .002$).
- Whether the tutor stated that a body needs to be chosen (20 cases) or not (12 cases, $p = .017$).

These features turn out to be strongly related. Table 3 shows why. The rows show different ways that the tutors prompted the goal of choosing a body. The columns show whether the teacher applied the principle, or whether the student made an error, got stuck, or applied the principle correctly. Notice that the 14 students who chose a body incorrectly are a subset of the 16 students whom the tutors asked to choose a body, who are in turn a subset of the 20 students who heard the tutor state that a body needs to be chosen. We believe it is really the first feature (errors) that makes a difference. That is, there were enough gainers in the smallest set to cause all three sets to be reliably associated with gains. This interpretation is consistent with the fact that by the time the knot principle comes up, the tutor and student had already discussed the need to

TABLE 2
Codes for the Knot Principle

Goal Setting Codes
- Did the tutor state that a body needed to be chosen?
- Did the tutor ask the student to select a body?

Initiation Codes
- Did tutor choose the body or did the student try to choose it?
- If the student tried to choose the body, did the student choose correctly, incorrectly or get stuck?

Explanation Codes
- Did the tutor point out that the knot is an object?
- If the student chose an incorrect body, did the tutor explain why the choice was wrong (e.g., by demonstrating that choosing a weight as a body leads to a dead-end)?
- Did the tutor explain why knots should be used as bodies in this problem (e.g., "You want to relate T3 and W1, and that knot is what you need.")?
- Did the tutor state the principle in general form?

TABLE 3
Relationship Between Goal Setting (Rows) and Initial Application of the Principle (Columns)

	T did	*Error*	*Stuck*	*Ok*	*Total*
T says body must be chosen					
T ask S to choose	0	14	0	2	16
T does not ask S to choose	3	0	1	0	4
T does not say body must be chosen					
T ask S to choose	0	0	0	0	0
T does not ask S to choose	8	0	1	3	12
Total	11	14	2	5	32

choose a body many times, so mentioning it one more time probably didn't make much difference. Thus, it is more likely that the gains were associated with making errors.

The Deceleration Principle

The deceleration principle is "If an object is slowing down, then it is accelerating in the direction opposite its movement." During instruction, this principle was first used in a problem where an elevator was slowing down as it descended. Students who knew the principle should have concluded that the elevator was accelerating upwards. On the tests, the students were asked to draw the acceleration of a truck that was slowing down while moving rightwards on a horizontal surface. Students who knew the principle should have drawn a leftward horizontal arrow.

The deceleration principle's learning opportunities. All 15 students who missed the principle on the pretest initially failed during training to correctly indicate that the elevator's acceleration was upward. In all cases, the tutors noticed the error and provided remedy. The tutors used several different tactics to teach the principle, including these:

- The tutor reasoned deductively from the definitions. In particular, the tutor usually began by asking the student for the definition of acceleration. The student should answer, "Change in velocity divided by the duration." The tutor next asked the student to draw the initial velocity of the elevator, the final velocity, and the change in velocity. The latter should be a short arrow pointing upwards. The tutor then asked the student the direction of the elevator's acceleration. The student should say "Up."
- The tutor reasoned by analogy. In particular, the tutor began by saying, "Suppose I am moving north. What direction would you have to push me in

order to slow me down?" The student should say, "South." The tutor then asked the student, "So according to Newton's law, what direction would my acceleration be?" The student should say, "South." The tutor then asked the student the direction of the elevator's acceleration. The student should say, "Up."

- The tutor used a Socratic approach. If the student said the acceleration is downward, the tutor asked what that would do to the velocity vector. The student should say that the velocity vector gets longer. The tutor asked what that would do to the elevator's speed. The student should say that the elevator would speed up. The tutor asked if the elevator is speeding up. The student should realize the contradiction and retract the belief that the elevator is accelerating downwards.
- The tutor gave some kind of mild negative feedback, such as "Are you sure the acceleration is downwards?" The student should then say something like, "No, I meant upwards."

Sometimes the tutors would try one tutorial tactic, then try a second if the first seemed not to work.

Features associated with learning the deceleration principle. The tactics (lines of reasoning) were coded by two coders ($\kappa = .88$). Codes are shown in Table 4. The only code that was reliably associated with gain was whether the

TABLE 4
Codes for the Deceleration Principle

Goal Setting Codes
- Did the tutor ask the student which direction the acceleration goes?

Initiation Codes
- None. All events were initiated by an error that was caught by the tutor.

Explanation Codes
- Which line of reasoning (tutorial tactic) was used?
- How many lines of reasoning were used?
- How many steps were in the line of reasoning, or how many steps were in all lines of reasoning if more than one was used?
- How many of the steps in the lines of reasoning were explicitly presented (tutors sometimes skipped steps)?
- How many of the steps in the lines of reasoning were produced by the tutor vs. by the student?
- How active was the student (number of steps produced by the student divided by total number of steps produced)?
- Was a general version of the principle stated?

Miscellaneous Codes
- Did the student draw a correct acceleration vector at the end or merely state that the acceleration was upward?

tutor stated a general version of the principle, namely "If a body is slowing down, the direction of its acceleration is opposite its motion" (p = .035; Coding was done by two coders with $\kappa = 1.0$). It was always the tutor who stated the generalization, never the student. Generalization facilitated correct answering of the posttest. Apparently, a correct answer to the vertical training situation (the elevator problem) was not entirely sufficient for the student to answer correctly in the horizontal testing situation (the truck problem). To increase generalization and transfer, it appears that the tutor should mention the critical concepts *slowing down* and *opposite*.

A Kinematics Equation

Several kinematics (time-rate-distance) equations are used in physics, and one of them is $s = v_0t + \frac{1}{2}at^2$, where s is the distance an object travels, t is the duration of travel, v_0 is the object's initial velocity and a is the object's acceleration. During training, this equation was used in a problem where a block starts at rest and slides down an inclined plane for 2 sec. It was tested by asking how far an object travels during 10 sec when starting from rest and accelerating at 5 m/s^2.

The kinematics equation's learning opportunities. The tutorial dialogs for this principle had the following general form. One of the students was able to produce a correct version of the equation. The other 22 students either could not produce any equation (9 cases) or produced incorrect equations (13 cases). If a student could not produce an equation, then the tutor did so and sometimes justified it by deriving it from the definitions of velocity and acceleration via either calculus (2 cases) or algebra (1 case). On the other hand, if the student produced an incorrect equation, the tutors responded in two ways:

- In 6 cases the tutor explained why the student's error was wrong. For instance, a common error was to use $s = vt$, where v is supposed to be the average velocity, but students used the final velocity instead. Tutors pointed this out and suggested using the target equation instead. In one case, the tutor also justified the target equation by deriving it via calculus.
- In 7 cases the tutor did not explain why the student's equation was wrong. For instance, when one student used $s = at^2$, the tutor simply pointed out that it should be $s = \frac{1}{2}at^2$. Additionally, in 4 cases, the tutor derived the equation via calculus.

Features associated with learning the kinematics equation. Table 5 shows the codes used. Two were associated with gains. First, if the student produced an incorrect equation and the tutor did not explain why it was wrong, then students rarely (in 1 of 7 cases) gained; but if the tutor explained why the

TABLE 5
Codes for the Kinematics Principle

Goal Setting Codes
- Did the tutor suggest finding a value for a variable?
- Did the tutor suggest writing an equation?
- Did the tutor suggest writing a kinematics equation?

Initiation Codes
- Whether the student produced a correct equation, an incorrect equation or no equation.

Explanation Codes
- Whether the kinematics equation was discussed before or after the value of acceleration was found (and thus, could be substituted into the equation).
- Whether the tutor asked the student to name or give values for the variables in the equation.
- Whether the student used the equation during training to calculate a numerical value for the distance.
- If the student produced an incorrect equation, did the tutor explain why it was wrong?
- Did the tutor derive the correct equation ($s = v_o t + {}^1/_2\, at^2$) via calculus or algebra?

Miscellaneous Codes
- Whether the student gave an incorrect answer on the pretest or gave no answer on the pretest.
- Whether the student made the same mistake they made on the pretest.

equation was wrong, then they usually (in 4 of 6 cases) gained, which was a marginally significant difference (p = .10). This finding suggests, contrary to claims by Sleeman et al. (1989), that tutor explanations of errors support learning, when merely correcting a mistaken equation is not sufficient to remedy the incorrect knowledge.

A second code was associated with gain. In none of the 8 cases in which the tutor derived the target equation via calculus or algebra did the students gain, whereas they gained in half of the 14 cases in which the tutor did not derive an equation mathematically (p = .02). Although we found that mathematical derivations of the equation were associated with low gains, Catrambone (1994b) found that such derivations improved gains. However, his test problems could be solved by applying just part of the derivation. In our case, the test problems used the same target equation as the training, so being able to apply only part of the derivation was not necessary for solving them correctly. Apparently, our tutors' derivations provided no useful new information and may only have confused the students.

The Compound Body Principle

Some physics problems are easier to solve if one treats two or more objects that move together as a single body. For instance, if the problem asks for the acceleration of a 40 kg boy on a 10 kg sled that is sliding down a hill, then it is easiest to treat the boy/sled combination as a single 50 kg body. The compound body principle is, "A set of objects that move together can be considered a single body." In the instruction, the principle is first used in a problem where two blocks, one sitting on top of the other, slide down a frictionless inclined plane. The principle is tested in a problem where two adjacent blocks sit on a horizontal frictionless plane, and a horizontal force is applied to the left side of the left block.

The compound body principle's learning opportunities. The tutorial dialogs had the following general form. Because the physics problem used in the training actually asked, "What is the acceleration of the two-block system," it strongly suggests that one should choose a compound body. Nonetheless, 4 students mistakenly chose a single block as the body. The other 21 students correctly chose the two blocks as the body but 5 showed uncertainty (e.g., by asking the tutor if it was correct). Regardless of how the body was chosen, in 23 of 25 cases the tutors confirmed that the two blocks should be treated as a single body, and in 9 cases even explained why (e.g., because they have the same acceleration or because they move together).

Features associated with learning the compound body principle. Table 6 shows the codes used. Of the 6 students who gained, all made a mistake or showed uncertainty, whereas of the 19 students who did not gain, 16 made the correct body choice without comment. The difference was significant ($p = .0005$).

TABLE 6
Codes for the Compound Body Principle

Goal Setting Codes
- Did the tutor ask the student to choose the body?

Initiation Codes
- Did the student choose the body incorrectly, correctly but showing uncertainty, or correctly and without comment?

Explanation Codes
- Did the tutor explain that the two blocks can be considered a single body because they move together?
- Did the tutor make any other explanations (e.g., there is no need to consider internal forces between blocks)?
- Did the tutor state the principle in a general form?

Miscellaneous Codes
- Did the student mention mass during the selection of the body (because it might be possible to work the problem by simply adding the masses of the two blocks together instead of conceptualizing the pair as a single body)?

No other significant differences were found among the other features that we coded for.

The Rotated Axes Principle

Although Newton's law is introduced as a vector equation, **F** = m**a**, where **F** and **a** are vectors, students are taught to apply it by choosing coordinate axes centered on the body, finding components of the vectors along each axis, and writing Newton's law in scalar form once for each axis (see Figure 1). The coordinate axes need not appear in their standard orientation, where the x-axis is horizontal and the y-axis is vertical. They can be rotated. Although different equations result from different rotations, the final answer will always be the same. Students can take advantage of this to simplify the equations. If a vector is perpendicular to an axis, then its component along that axis is zero, so it need not appear in the equation for that axis. The axes can be rotated to make certain vectors perpendicular to the axis, and thus simplify the equations.

The general version of the axis rotation principle is to choose axes that simplify the equations. However, this version of the principle is not operational. One would have to mentally envision the equations that would result from different rotations of the axes to determine which rotation creates the simplest equations. Not even experts could do that easily. Thus, there are two more specific versions of the principle that are commonly used:

- If the sought quantity is the magnitude of a vector, such as an acceleration or a force, then rotate the axes so that one axis is aligned with it. This means that the unknown will appear in only one equation.
- If the sought quantity is a scalar, such as mass, or there is no sought quantity, then rotate the axes so that the maximum number of vectors lie along axes. If a vector lies along an axis, its magnitude will appear in only one equation.

During training, the rotated axis principle was used on three problems (P2, P3, and P5). In all three cases, the first version of the principle was appropriate because the sought quantities were vectors. Tutors' explanations usually stated that the axes should be rotated to align with the direction of the unknown quantity. Sometimes the tutors would also state that rotating the axes simplifies the equations.

Unfortunately, the test problem required much further transfer than we had intended. The problem provided several vectors without specifying which one was sought, and asked the student to draw appropriate coordinate axes. Thus, it could be solved by the second version of the rotated axes principle. Yet that version was not relevant during the tutoring sessions, so there was no tutoring on it.

The rotated axes principle's learning opportunities. During the first problem where the rotated axis principle could be used, the tutoring proceeded as follows. In 12 cases, the tutor suggested rotating the axes before giving the student a chance to draw them. In the remaining 6 cases, the student chose the axes and usually (4 cases) chose nonrotated ones. However, even when the tutor did suggest rotating the axes before giving the student a chance to draw them, 1 of the 12 students apparently misunderstood, because the student failed to rotate the axes even after receiving the suggestion. On the other hand, if they chose correctly, they did not express uncertainty. Regardless of how the axes were chosen, the tutors often (in 13 of 18 cases) mentioned the first version of the principle ("It's generally a good idea to put one axis along the unknown."). In 4 of those 13 cases, the tutors continued by mentioning the general version ("That will simplify the equations."). There were no cases where the tutors gave the general explanation without giving the specific explanation.

On the second problem where the rotated axis principle could be used, the tutor suggested rotating the axes before the student had a chance in only 3 cases. In the third problem, the tutor usurped the student's prerogative in only 1 case. Thus, the tutors were fading the scaffolding, which has been observed before in tutoring (McArthur et al., 1990, Figure 3).

The tutors also gave explanations somewhat less frequently on subsequent problems. On both the second and third problems, they gave explanations in 10 of 18 cases, whereas they gave explanations in 13 of 18 cases on the first problem.

Features associated with learning the rotated axes principle. Table 7 shows the codes used. None were associated with gain. In particular, for each of the three problems, we tested whether any code for that problem was associated with gain. We also formed codes that aggregated across problems, such as "Did the student make an error on any problem," and tested their association with gain.

TABLE 7
Codes Used for the Rotated Axes Principle

Goal Setting Codes
- Did the tutor mention that the axis can or should be rotated?

Initiation Codes
- Did the tutor suggest how much to rotate the axes before giving the student a chance?
- Did the student draw rotated axes or nonrotated ones?

Explanation Codes
- Did the tutor mention the specific version of the principle?
- Did the tutor mention the general version of the principle?
- Did the tutor mention both versions of the principle?

Miscellaneous Codes
- If the student drew nonrotated axes, did the tutor explain either the specific or general principle?

TABLE 8
Association of Errors on the Rotated Axis Principle With Subsequent Error

	S erred on the second problem?	
S erred on the first problem?	*No*	*Yes*
Yes	5	0
No	7	3

Again, no significant associations were found. This is not too surprising, given that the training was on one specific version of the principle and the testing was on the other version.

Because we had unintentionally used a far transfer problem on the test, we tried in vain to find evidence of learning by seeing if tutoring reduced errors during training. We used the codes of Table 7 as independent variables, but took errors on the second and third problem as the dependent variable. For instance, Table 8 shows the results for whether or not making an error on the first problem is associated with making an error on the second problem. The table indicates that if the student erred on the first problem (and received remedy), then the student was less likely to error on the second problem. This is the trend one would expect, but it did not reach significance. Note that only 15 of the 18 students could be used in this analysis because for the other 3, the tutor usurped their chance of making an error on the second problem. The use of subsequent errors as a dependent variable caused low *N* on all of the tests of association, which is probably why none of them reached significance.

THIRD ANALYSIS: GENERALIZATION OVER PRINCIPLES

Table 9 summarizes the results from the preceding analysis. As expected, for different principles, different features were associated with gain. This is why we did not observe strong associations in the first analysis. However, it leaves one to wonder what can be said about learning in general regardless of the principle being learned. To address that, we conducted a third analysis. It is motivated by the findings from the second analysis. It used the same coding as the second analysis, but aggregated the codes across principles in ways that were motivated by patterns observed in the second analysis and by the theoretical issues discussed in the introduction.

Are Impasses Associated With Learning?

One of our main questions concerned the effect of impasses on learning. With the knot principle and the compound body principle, impasses were associated with

TABLE 9
Codes Associated With Gains

Knot Principle	
	• The student chose an incorrect body (vs. the choosing the correct body or letting the tutor choose the body).
Deceleration Principle	
	• The tutor stated a general version of the principle (vs. neither tutor nor student stating a general version of the principle).
A Kinematics Equation	
	• If the student produced an incorrect equation, the tutor explained why it was wrong (vs. not explaining why it was wrong).
	• The tutor did not use mathematical derivations of the equation (vs. the tutor used an algebraic or calculus derivation of the equation).
Compound Body Principle	
	• The student chose an incorrect body or chose the correct body but expressed uncertainty (vs. choosing the correct body without expressing uncertainty).
Rotated Axes Principle	
	• None (probably because the test problem could not be solved with the principle taught during the tutoring).

gains. There appear to be several reasons why no such association was observed for the other three principles:

- For the rotated axes principle, the test problem did not measure acquisition of the principle taught during training, so unfortunately, data from this principle are worthless for testing the impasse hypothesis or any other hypothesis. In fact, they will be ignored in subsequent discussions.
- For the deceleration principle, all the learning opportunities began with an impasse.
- For the kinematics equation, 22 of the 23 learning opportunities began with an impasse.

The useable data suggest that impasses increase the likelihood of learning but do not guarantee it. This is evident in Table 10, which presents an analysis of the learning opportunities from four of the five principles (i.e., data from the rotated axes principle are not included). The analysis involved aggregating the coding from the first analysis across principles according to the definition of impasses as episodes where the student got stuck, made an error, or did the correct action while showing signs of uncertainty.

The association apparent in Table 10 is statistically reliable, $\chi^2(94) = 13.4$, $p = .001$, and shows that learning opportunities that include an impasse result in gains about half the time, whereas gains are less frequent when learning

TABLE 10
Impasses are Associated With Gains

	Gain	*No-gain*
Impasse: Student gets stuck, errs, or does correct action with uncertainty	33	29
No impasse: Tutor does correct action	3	11
No impasse: Student does correct action without signs of uncertainty	2	17

opportunities have the tutor doing the action (21%) or the student doing the action correctly without signs of uncertainty (11%). When students did the correct action without signs of uncertainty despite having failed to do so on the pretest, it was mostly on the compound body principle (16 of 22 cases), where the problem statement strongly implied that the student should use a compound body. Thus, it seems that impasses are indeed associated with gains because when they are absent, students seldom learn.

Are Explanations More Effective in the Context of an Impasse?

As Table 9 indicates, explanations, albeit of different kinds, were associated with learning for the Deceleration principle and for the Kinematics equation. Why were explanations not associated with learning for the Compound body principle or the Knot principle? Almost all the learning opportunities of the Deceleration and Kinematics principles included impasses, but impasses occurred on only some of the learning opportunities for the Compound body and Knot principles. This suggests that explanations may be more effective in the context of an impasse.

To test this, the statistical tests for association were redone for the Compound body and Knot principles including only learning opportunities where impasses occurred. None of the explanation codes were reliably associated with gain. Although this could be owing to lack of statistical power, trends were not apparent in the data (see Tables 11 and 12; the cells indicate the number of gains over the total number of learning opportunities).

This suggests that it may be something about the content of the Knot and Compound Body principles that makes explanations ineffective. They do share the common feature that they advise the student which object to pick as the body. This is often quite salient visually, so it may be that some kind of shallow, visual learning suffices for transfer to the posttest. Even if providing explanations encourages deeper learning, the deeper learning may be no better than shallow learning at causing transfer to our posttest.

TABLE 11
For the Knot Principle, Are Explanations in the Context of Impasses Associated With Gain?

	Explain	*No explain*	*p(Fisher)*
Tutor points out that the knot is an object	10/13 (77%)	3/3 (100%)	1.0
Tutor explains why the student's choice is wrong	8/11 (73%)	5/5 (100%)	0.509
Tutor explains the general principle	3/5 (60%)	10/11 (91%)	0.214
Tutor explains why knot is right for this problem	8/8 (100%)	5/8 (63%)	0.200

TABLE 12
For the Compound Body principle, Are Explanations in the Context of Impasses Associated With Gain?

	Explain	*No Explain*	*p(Fisher)*
Tutor explains: They move together so consider as one	2/3 (67%)	4/6 (67%)	1.0
Tutor explains the general principle	0/0	6/9 (67%)	N/A
Tutor gives any other explanation	2/4 (50%)	4/5 (80%)	0.524

Is Mentioning the Goal Effective in the Context of Impasses?

As discussed in the introduction, mentioning goals ought to be associated with learning at least in some cases. In the earlier analysis, discussion of goals was associated with learning for the Knot principle but not for any of the other three principles. Even for the knot principle, it appears that the association is the result of a confound with impasses, as argued in connection with Table 3. Just to be thorough, we reanalyzed the data for the Compound Body and Knot principles, including only learning opportunities where there were impasses. Mentioning goals was not associated with gain.

GENERAL DISCUSSION

Here we first consider the two main objectives of the study, then its for human and computer-based tutoring.

Objective 1: Understanding Coached Problem Solving

One objective was to add detail to the emerging picture of coached problem solving, and in particular, to help us envision the learning opportunities for principles that occur during coached problem solving. In reporting our first analysis, we described the learning opportunities associated with each of the five principles. In this section, we point out some of the similarities among them.

Learning opportunities can be viewed as consisting of three unordered parts. During the goal setting part, the tutor may prompt the student to apply the principle by telling the student what the current goal is. This occurred in 61% of the learning opportunities (see Table 13), counting only the initial application of the Rotated Axes principle. This is a surprisingly high frequency, given that McArthur et al. (1990) found goals mentioned much less frequently in their algebra problem solving transcripts. It may be that their participants were more familiar with the relevant goal management knowledge than our participants, and thus needed less prompting.

The second part, initial principle application, consisted of either the student or the tutor applying the principle. In our corpus, it was usually the student who applied the principle. The tutor applied the principle in only 21 (19%) of the 113 learning opportunities (counting only the initial learning opportunities for the rotated axis principle). Students reached an impasse on 67 (59%) of the 113 learning opportunities. This is not surprising given that we studied only learning opportunities where the student missed the principle on the pretest. If the student made an error, our tutors always pointed it out if the students did not detect it already, which is consistent with earlier studies of coached problem solving (e.g., Merrill et al., 1995).

The third part consists of explanations and other discussion of the principle application. The tutor produced most of the explanations. We had hoped to see cases where the student provided most of the explanation, but this seldom occurred. For the Deceleration and Kinematics principles, some of the explanations consisted of elaborate derivations of the principle. These were never positively associated with gain, and mathematical explanations were inversely associated with gains for the Kinematics principle. For the other three principles, explanations generally did not involve elaborate derivations, possibly because there are none for these principles. Moreover, explanations of these principles were not associated with gain.

TABLE 13
How Frequently Did the Tutor Mention the Goal?

Principle	*Frequency*
Knot	20/32 (63%)
Rotated axis	8/18 (44%)
Deceleration	13/15 (87%)
Kinematics equation	16/23 (70%)
Compound body	10/25 (40%)
Average	61%

Although the three parts often occurred in the order of goal prompt, initial application and explanation, they did not always occur in that order. For instance, sometimes the student would make an attempt to apply the principle, then the tutor would prompt by mentioning the goal. Sometimes tutors included a little bit of the explanation (e.g., "Notice that the knot is also an object.") as a prompt even before the student made a first attempt at the goal.

In short, the picture to take home of a typical learning opportunity in this study is: The tutor usually suggests a goal; the student usually tries to apply the principle and often reaches an impasse; and the tutor gives most of the subsequent explanation. To be frank, there were few cases of the kind of inspired tutoring observed by Collins and Stevens (1982), despite the fact that our tutors had years of experience, had great enthusiasm for tutoring, and knew that their tutoring was being recorded and analyzed.

Objective 2: Tutoring Features Associated With Learning

The second objective of this research was to find features of the learning opportunities that were associated with learning. These proved surprisingly elusive. When we analyzed all the principles together, we found only that shorter learning opportunities were associated with more frequent gains. It could be that shorter explanations are more effective, but it could also be that easily-learned principles elicit shorter discussions.

When we controlled for principle learnability by analyzing each of five principles separately, we did observe correlates of learning. However, different features were associated with the learning of different principles. The knot principle and the compound body principle were learned when students reached impasses, as predicted by several earlier studies (Brown & VanLehn, 1980; Carroll & Kay, 1988; Siegler & Jenkins, 1989; VanLehn, 1987, 1990, 1999; VanLehn & Jones, 1993b; VanLehn et al., 1992). For the deceleration and the kinematics principles, almost all the students reached impasses, so only explanations or goal-prompting could be associated with learning. It turned out that generalization was important for deceleration, which is consistent with Catrambone's (1994a) findings. Contrary to Sleeman et al. (1989), explanation of why the student's equation was wrong may have been important for learning the correct kinematics equation. Contrary to Catrambone (1994b), mathematical derivations of the correct kinematics equation actually impeded learning.

This diversity of findings prompted a third analysis of the data. It found that learning was more common at impasses than when the tutor did the action correctly or when the student did the action without signs of an impasse. This suggests that letting the student try to do an action even if they will get stuck or make an error is better than showing them how to do it.

Dweck (1986) and Dweck and Leggett (1988) found that changing reward structures could change students from a performance orientation to a learning orientation. In our studies, when students reach an impasse, they discover that they need to learn something, so they may adopt a learning orientation at least for the duration of the learning opportunity. If so, then one would expect explanations and goal prompting to be more effective in the context of an impasse than they would otherwise be. That did not turn out to be the case, although the null results could be because of low statistical power.

Implications

When all is said and done, what has been learned from this close examination of learning opportunities during coached problem solving? First, impasses appear to be strongly associated with learning. This suggests that tutors should encourage impasses. Our tutors always gave feedback on errors, unless the students detected the errors first, but they should perhaps go even further. Of the 33 cases just mentioned where there were no impasses, there were 11 cases when tutors applied the principle without even giving the student a chance first. Although it is possible that the tutors deliberately avoided the impasses for affective reasons, such as preventing the students from losing face, it may also be that the tutors overestimated the students' competence or were more enthusiastic about making progress through the problem than letting the student make mistakes and learning from them. As a general policy, tutors should let impasses occur unless there are compelling reasons (e.g., affective ones) for avoiding them. The other 22 cases of non-impasses occurred mostly when students guessed correctly about the compound body due to the way the problem was worded. These cases could perhaps be avoided by rewording the problem.

The large number of null results for explanations and goal prompting suggest that tutorial explanations and other tactics are perhaps much less important than we initially thought. McKendree (1990) showed that a tutoring system that merely flags errors yielded significantly less learning than when it gave hints as well. However, most of the benefit seemed to be due to the students learning goal management skills, as opposed to principles. The analyses presented here suggest that tutorial explanations are associated with gains in only a few cases.

If explanations are not associated with learning, then we face the counterintuitive but intriguing possibility that the content of the tutor's comments may not matter much. The main effect of tutorial explanations may be to prompt students to think harder with the knowledge that they already possess. If so, then it doesn't matter what the tutors say or even whether the students understand it; the mere fact that the tutors are talking about a principle motivates the students to think harder about it. This would explain why individual explanation features often did not correlate with learning.

This is also consistent with Chi's recent studies showing that when tutors are prohibited from generating explanations and may only give content-free prompts to students, then students learn just as much as when tutors are allowed to tutor normally (Chi et al., 2001). However, Chi's participants were reading a textbook instead of solving a problem, so her participants had access to authoritative knowledge, whereas our problem-solving students never referred to the text while working with the tutor, so their only authoritative source was their tutor. It is not clear whether zero-content prompting will work as well with coached problem solving as it does with coached reading. Moreover, in the prompting condition of Chi's studies, the prompters may have kept prompting the student until the student had produced an explanation that satisfied the prompter. Nonetheless, there seems to be a growing suspicion that human tutorial explanations may not be all that helpful for students. Tutorial behavior that gets students to think, such as generating opportunities for impasses or giving zero-content prompts, may be the key to why tutoring is so effective. If a feedback loop is added that keeps students thinking and talking until they have produced an explanation that satisfies the tutor, then that might further increase the effectiveness of such explanation-free tutoring.

On the other hand, zero-content prompting probably fails whenever the students simply cannot generate a sufficient explanation by themselves. Renkl (2001) has shown that learning can be improved by providing tutorial explanations whenever students fail to generate a correct self-explanation.

This suggests that an optimal tutoring strategy may be to (a) let the student reach an impasse, (b) prompt them to find the right step and explain it, and (c) provide an explanation only if they have tried and failed to provide their own explanation. Human tutors often fail to do step (b) and sometimes even fail to do step (a). On the other hand, some intelligent tutoring systems approximate this tutoring strategy in that they always let students try to do a step before giving hints on it, and their hint sequences gradually increase in content. However, they do not elicit an explanation from the students, nor would they be able to understand it if they did, so they do not know when the students' explanation is satisfactory. This suggests that the next major advance in tutoring technology might be tutors that elicit and understand student explanations (e.g., Rose et al., 2001; VanLehn et al., 2002).

Given that a tutorial explanation is finally necessary, what kind of explanation should be given? It turned out in this study that explanations that were just deep enough to allow students to solve the posttest problems were more effective than deeper explanations. For instance, on the Deceleration principle, tutors provided several interesting explanations for the principle, and sometimes even got students involved in producing parts of the explanation. One would think that this should cause deep learning of the principle. However, it turned out that gains were associated only with stating the principle in a general form ("If an object is moving in a straight line and slowing down, then its acceleration is in the opposite direction from its motion."). Because the posttest problem showed an object moving horizontally and slowing down, while the training problem showed an object moving vertically and slowing down, the students only needed to incorporate the key concepts *slowing down* and *opposite direction* into their knowledge to transfer their training experience to the posttest. Any explanation that went beyond this was simply wasted breath, according to one interpretation of our results. Similarly, the mathematical derivations of the Kinematics principle didn't help; in fact, they were associated with less gain. The explanations of the two body-choice principles, Knot and Compound Body, had no apparent benefits for students (even after an impasse, which might have motivated them to attend to the explanations), probably because a shallow, visual form of learning would suffice to transfer the training experience to the posttest. Lastly, in the initial analysis that aggregated learning opportunities over principles, the only feature that was associated with gains was number of words, which suggests that shorter explanations are more effective. These results suggest that when tutors must give an explanation because the student cannot, the tutor's explanation should be as simple and short as possible. More elaborate explanations, while perhaps beneficial in non-tutorial settings, appear to have no benefits in this study.

To put it in motto form, tutors should "ask more and tell less." Although our results are consistent with this motto, considerably more research is needed before we really understand how to optimize human tutoring and computer-based tutoring.

ACKNOWLEDGMENTS

This research was supported by the Cognitive Science Division of ONR under contracts N00014-94-1-0674, N00014-95-1-0950 and N00014-96-1-0260.

REFERENCES

Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent tutoring systems: 5th international conference, ITS 2000* (pp. 292–303). Berlin: Springer.

Anderson, J. R. (1982). The acquisition of cognitive skill. *Psychological Review, 89,* 369–406.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science, 14*(4), 467–505.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167–207.

Anderson, J. R., Farrell, R., & Saurers, R. (1985). Learning to program in Lisp. *Cognitive Science, 8,* 87–129.

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

Brown, J. S., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science, 4,* 379–426.

Bundy, A., Byrd, L., Luger, G., Mellish, C., & Palmer, M. (1979). Solving mechanics problems using meta-level inference. In *Proceedings of the Sixth International Joint Conference on AI* (pp. 1017–1027). San Mateo, CA: Kaufmann.

Carroll, J. M., & Kay, D. S. (1988). Prompting, feedback and error correction in the design of a scenario machine. *International Journal of Man-Machine Studies, 28,* 11–27.

Catrambone, R. (1994a). The effects of labels in examples on problem-solving transfer. In A. Ram & K. Eiselt (Eds.), *Proceedings of the sixteenth annual conference of the Cognitive Science Society* (pp. 159–164). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Catrambone, R. (1994b). Improving examples to improve transfer to novel problems. *Memory and Cognition, 22*(5), 606–615.

Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10,* S33–S49.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 15,* 145–182.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.

Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25,* 471–533.

Cho, B.-I., Michael, J. A., Rovick, A. A., & Evens, M. W. (2000). An analysis of multiple tutoring protocols. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent tutoring systems: 5th international conference* (pp. 212–221). Berlin: Springer.

Collins, A., & Stevens, A. (1982). Goals and strategies for inquiry teachers. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 65–119). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Corbett, A., McLaughlin, M., & Scarpinatto, K. C. (2000). Modeling student knowledge: Cognitive tutors in high school and college. *User Modeling and User-Adapted Interactions, 10,* 81–108.

de Kleer, J. (1975). *Qualitative and quantitative knowledge in classical mechanics* (Tech. Rep. No. AI-TR-352). Cambridge, MA: MIT AI Laboratory.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41,* 1040–1048.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273.

Elio, R., & Scharf, P. B. (1990). Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization. *Cognitive Science, 14,* 579–639.

Fox, B. A. (1993). *The human tutorial dialogue project: Issues in the design of instructional systems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Frederiksen, C. H., Roy, M., & Bedard, D. *Expert tutoring and learning: Analysis of tutorial dialogue in engineering*. Manuscript submitted for publication.

Graesser, A. C., Person, N., & Magliano, J. (1995). Collaborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology, 9,* 359–387.

Heffernan, N. T. (2001). *Intelligent tutoring systems have forgotten the tutor: Adding a cognitive model of human tutors*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-on-one tutoring. *Journal of the Learning Sciences, 5*(1), 23–49.

Jones, R. M., & VanLehn, K. (1994). Acquisition of children's addition strategies: A model of impasse-free, knowledge-level learning. *Machine Learning, 16*(1/2), 11–36.

Larkin, J. (1983). The role of problem representation in physics. In D. Gentner & A. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Larkin, J. H. (1981). Enriching formal knowledge: A model for learning to solve textbook physics problems. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. *Cognitive Science, 4,* 317–345.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J.-L. (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. 75–105). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lewis, M. (1989). *Developing and evaluating the CMU algebra tutor: Tension between theoretically driven and pragmatically driven design*. Paper presented at the American Educational Research Association, San Franscisco, CA.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction, 7*(3), 197–244.

McDermott, J., & Larkin, J. H. (1978). Re-representing textbook physics problems. In *Proceedings of the second national conference of the Canadian Society for Computational Studies of Intelligence*. Toronto, Ontario.

McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human-Computer Interaction, 5,* 381–413.

Merrill, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction, 13*(3), 315–372.

Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences, 2*(3), 277–306.

Moore, J. D. (1993). What makes human explanations effective? In W. Kintsch (Ed.), *Proceedings of the fifteenth annual conference of the Cognitive Science Society* (pp. 131–136). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–56). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Novak, G. S., & Araya, A. A. (1980). Research on expert problem solving in physics. In *Proceedings of the first national conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press.

Pollack, M. E. (1990). Plans as complex mental attitudes. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communications* (pp. 77–104). Cambridge, MA: MIT Press.

Porayska-Pomsta, K., Mellish, C., & Pain, H. (2000). Aspects of speech act categorisation: Towards generating teachers' language. *International Journal of Artificial Intelligence in Education, 11,* 254–272.

Priest, A. G., & Lindsay, R. O. (1992). New light on novice-expert differences in physics problem solving. *British Journal of Psychology, 83,* 389–405.

Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal, 24*(1), 13–48.

Reimann, P., & Schult, T. J. (1993). Understanding and using worked-out examples: A computational model. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge*. Amsterdam: Elsevier.

Reimann, P., Wichmann, S., & Schult, T., J. (1993). A learning strategy model for worked-out examples. In P. Brna, S. Ohlsson, & H. Pain (Eds.), *Artificial intelligence in education: Proceedings of AI-ED93* (pp. 290–297). Charlottesville, VA: Association for Advancement of Computing in Education.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction, 12*(5), 529–556.

Rose, C. P., Jordan, P. W., Ringenberg, M., Siler, S., Vanlehn, K., & Weinstein, A. (2001). Interactive conceptual tutoring in Atlas-Andes. In J. D. Moore, C. Redfield, & W. L. Johnson (Eds.), *Artificial intelligence in education: AI-Ed in the wired and wireless future* (pp. 256–266). Washington, DC: IOS.

Schoenfeld, A. H., Gamoran, M., Kessel, C., & Leonard, M. (1992). Toward a comprehensive model of human tutoring in complex subject matter domains. *Journal of Mathematical Behavior, 11*(4), 293–319.

Shute, V. J., & Psotka, J. (1996). Intelligent tutoring systems: Past, Present and Future. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology*. Macmillan Library Reference.

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sleeman, D., Kelly, A. E., Martinak, R., Ward, R. D., & Moore, J. L. (1989). Studies of diagnosis and remediation with high school algebra students. *Cognitive Science, 13,* 551–568.

VanLehn, K. (1987). Learning one subprocedure per lesson. *Artificial Intelligence, 31,* 1–40.

VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.

VanLehn, K. (1991). Rule acquisition events in the discovery of problem solving strategies. *Cognitive Science, 15*(1), 1–47.

VanLehn, K. (1996). Cognitive skill acquisition. In J. Spence, J. Darly, & D. J. Foss (Eds.), *Annual review of psychology* (Vol. 47, pp. 513–539). Palo Alto, CA: Annual Reviews.

VanLehn, K. (1999). Rule learning events in the acquisition of a complex skill: An evaluation of Cascade. *Journal of the Learning Sciences, 8*(2), 179–221.

VanLehn, K., & Jones, R. M. (1993a). Integration of analogical search control and explanation-based learning of correctness. In S. Minton (Ed.), *Machine learning methods for planning* (pp. 273–315). Los Altos, CA: Morgan Kaufmann.

VanLehn, K., & Jones, R. M. (1993b). Learning by explaining examples to oneself: A computational model. In S. Chipman & A. Meyrowitz (Eds.), *Cognitive models of complex learning* (pp. 25–82). Boston, MA: Kluwer Academic.

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

Vanlehn, K., Jordan, P. W., Rose, C. P., Bhembe, D., Bottner, M., Gaydos, A., et al. (2002). The architecture of Why2-Atlas: A coach for qualitative physics essay writing. In S. A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), *Intelligent tutoring systems, 2002, 6th international conference* (pp. 158–167). Berlin: Springer.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17,* 89–100.