To appear in M. A. Gernsbacher & S. Derry (Eds.) *Proceedings of the Twenth-first Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum, 1998.

What Makes a Tutorial Event Effective?

Kurt VanLehn (VanLehn@cs.pitt.edu)
Stephanie Siler (siler+@pitt.edu)
Charles Murray (rmurray+@pitt.edu)
Learning Research and Development Center
University of Pittsburgh, Pittsburgh PA 15260
William B. Baggett (baggett@vms.cis.pitt.edu)
School of Computer Technology; 107 6th Street
Pittsburgh, PA 15222

Abstract

Although tutoring by expert human tutors is usually effective, it is not always. By contrasting cases where tutoring does and does not result in learning, we can find out what causes learning during tutoring. Approximately 125 hours of physics tutorial dialog were analyzed to see what features of the dialog were associated with learning. Successful learning appears to require that the student make an error or reach an impasse; too much help can prevent learning. Features of successful tutorial explanations appear to be different for different pieces of knowledge. For instance, some pieces of knowledge are learned only if the tutor emphasizes generalization, whereas other learning requires that the tutor first explain why the student's error is wrong.

Introduction

Tutoring by an expert human tutor is one of the most effective forms of instruction known (Bloom, 1984). However, not all tutoring is equally effective. The research presented here tries to find out what distinguishes effective tutoring from ineffective tutoring. The results have implications not only for theories of learning, but also for improving both human and computer tutoring.

Methods

A study was conducted in which 2 expert physics tutors worked with 42 college students for approximately 3 hours each. The students were tested before and after the tutoring session in order to determine what they learned. Tutoring sessions were recorded and selected recordings were transcribed for analysis.

The 5 physics problems covered in the tutoring sessions were analyzed in order to find out what rules the students could learn. We use "rule" to stand for any piece of physics knowledge, both procedural and conceptual. For instance, one rule is, "If a taut string is attached to an object, there is a tension force on the object exerted by the string." The physics rules were the ones used in Andes (VanLehn, 1996), Pola (Conati & VanLehn, 1995), Olae (Martin & VanLehn, 1995) and Cascade (VanLehn, Jones, & Chi, 1992).

A test was designed to assess the learning of each rule used during the training problems. The same test was used for both pre- and post-testing.

For each student and rule, each test was scored according to whether the student used the rule or not (2 coders, with an interrater reliability of .95). If a subject failed to use a rule on the pre-test but used it on the post-test, the subject was said to "gain" that rule. If a subject failed to use a rule on both the pre- and post-tests, then the subject was said to "not gain" that rule.

Since we could not feasibly analyze all 42 students, we chose a subset of 8 that would maximize the contrast among learning and non-learning. For each of the 2 tutors, we selected for analysis the 2 students with the most gains and the 2 students with the fewest gains.

For each of the 8 subjects, potential learning events were located in the protocols. A potential learning event is an episode where the student has the opportunity to learn one of the rules that the student missed on the pre-test. More specifically, an episode was classified as a potential learning event if the tutor and student were either discussing the rule or applying the rule in order to solve the problem.

General Features

Our first analysis sought to find general features of potential learning events that would explain why some caused learning and some did not. We coded the potential learning events using the following codes, then correlated them with the gains. (Pearson correlations are shown in parentheses when $p<.05$).

- Who initiated the discussion: student or tutor?
- What initiated the discussion: a student–flagged error, a tutor–flagged error ($r = -.24$), or the student getting stuck?
- How many of the key ideas behind the rule 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 rule: the student (r = .34), the tutor, 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 (r = –.24)?
- How many words were uttered by participants during the discussion (r = –.40)?
- How many impasses or errors (r =–.29) occurred for this rule overall?

The coding was done by two coders. The reliabilities were calculated separately for each code, and all were judged acceptable.

The pattern of correlations is consistent with the idea that some of the rules were harder to learn than others—they required more words, caused more errors, and the errors were not caught by the student. This suggested looking at the correlations with the difficulty of the rule partialed out. We used the number of key ideas underlying the rule as a measure of the rule's difficulty. When difficulty was partialed out, only two features were significantly correlated with gain:

- The number of words uttered by the participants during discussion of the rule (r = –.27).
- The number of times the correct conclusion was first generated by the student (r = .30).

However, neither correlation explained much of the variance in learning. This suggested that there were other factors besides these general features that determined when a student would learn from a tutoring event.

Rule-Specific Features

We suspected that the reason we could not observe strong explanatory patterns was that our analysis sought a single pattern that explained every rule's learning. Perhaps different kinds of tutorial interactions were important for different rules. Thus, our next analysis examined each rule separately, looking for features of the tutorial dialogs that explained just that rule's learning.

Because we needed enough tutorial events per rule to get statistical power, we abandoned the labor saving device of examining only 8 subjects and analyzed all 42 subjects. However, we did not examine all rules. We examined a rule only if 5 or more subjects gained it and 5 or more subjects failed to gain it. These constraints were necessary in order to have enough variance to explain. The constraints eliminated all but 5 rules.

We first checked whether gains could be explained by the overall competence of the students. For instance, for the first rule discussed below, of the 15 students who missed the rule on the pre-test, 8 gained and 7 did not, and the pre-test scores of the gainers (25.4) were not significantly different (p=.57) from the pre-test scores of the non-gainers (27.1). In fact, for none of the 5 rules were the pre-test scores of the gainers significantly different from pre-test scores of the non-gainers.

For each rule, we separately checked whether the competence of the tutors could explain the gains. For none of the 5 rules were gains associated with tutor (Chi-square test).

In short, it appears that learning is associated with what occurred during the tutorial dialog, and not who was involved. In each of the sections below, the tutorial dialog features associated with one rule's gains are described.

The Deceleration Rule

The deceleration rule is "If an object is slowing down, then it is accelerating in the direction opposite its movement." This rule was taught in the context of an elevator that is slowing down as it descends. Students who knew the rule would conclude that the elevator is accelerating upwards. On the tests, the students are asked to draw the acceleration of a truck that is slowing down while moving rightwards on a horizontal surface. Students who knew the rule would draw a leftward horizontal arrow.

All 15 students who missed the rule 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 remediation. The tutors used 8 different tactics to teach the rule, including:

- The tutor begins by asking the student for the definition of acceleration, which is, "change in velocity divided by the duration." The tutor next asks 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 asks the student what direction the acceleration of the elevator is. The student should say "Up."
- The tutor poses an analogy 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 asks the student, "So according to Newton's law, what direction would my acceleration be?" The student should say, "South." The tutor then asks the student what direction the acceleration of the elevator is. The student should say, "Up."
- The tutor uses a Socratic approach. If the student says the acceleration is downward, the tutor asks what that would do to the velocity vector. The student should say that the velocity vector gets longer. The tutor asks what that would do to the elevator's speed. The student should say that the elevator would speed up. The tutor asks if the elevator is speeding up. The student should realize the contradiction and retract the belief that the elevator is accelerating downwards.
- The tutor gives some kind of mild negative feedback, such as "Are you sure the acceleration is downwards?" The student then says something like, "No, I meant upwards."

Sometimes the tutors would try one line of reasoning, then try a second if the first seemed not to work.

The 8 tactics (lines of reasoning) were coded by two coders, with inter-rater reliability of .88. The following features of the tutorial dialog were found *not* to be associated with gain:

- 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 the student was (number of steps produced by the student divided by total number of steps produced).
- Whether the student drew a correct acceleration vector at the end or merely stated that the acceleration was upward.

In other words, it did not seem to be the lines of reasoning or even their quality that determined gain.

However, what was reliably associated with gain was whether the tutor stated a general version of the rule, namely "If a body is slowing down (decelerating), the direction of its acceleration is opposite its motion" (χ^2=12.4, p=.0004; Coding was done by two coders with reliability of 1.0.) A correct answer to the vertical training situation (the elevator problem) was not sufficient for the student to answer correctly in horizontal testing situation (the truck problem). In order to obtain generalization and transfer, the tutor had to mention the critical concepts "slowing down" and "opposite."

The Knot Rule

Students often think that the only objects that they should apply Newton's law to 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 that one will apply Newton's law to.) A common massless object is a knot formed by tying together several massless strings. Ideally, the rule 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 rule, "A knot can be a body."

In the training, this rule was used on a problem where two blocks are hung from a harness of 5 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 rule was used on a problem where two men are pulling a cart with a harness that has 3 strings and one knot. Thus, students must transfer the application of the rule from a vertical case to a horizontal case, and from a more complex harness to a simpler one.

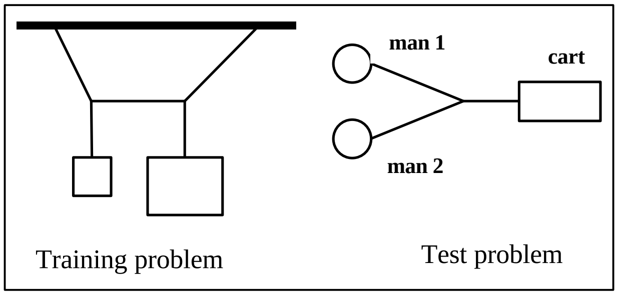

Figure 1

Tutoring on the knot rule proceeded as follows. Tutors sometimes mentioned quite early in the problem that there was a knot at the junction 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 needs to be chosen, they either explicitly stated that a body must be chosen (20 cases) or did not explicitly state this (16 cases). If they did not state that a body must be chosen, then they just started using knots as bodies, for instance, by asking the student to draw the forces acting on one of the knots. If the tutor stated that a body must be chosen, then sometimes the tutors chose the knot themselves (4 cases) and sometimes they asked the student to choose the body (16 cases). When the students chose the body, the student usually chose incorrectly (14 cases). Regardless of who chose the knot as the body, the tutor would sometimes explain the rule (13 cases). During those explanations, they would sometimes (6 cases) state a general version of the rule, such as "A body should be chosen that connects objects with known properties to objects whose properties we seek."

The following features were associated with gain:

- Whether a student incorrectly chose the body (14 cases) or not (22 cases). (χ^2=6.5, p=.011)
- Whether a student chose the body (16 cases) or not (20 cases). (χ^2=10.2, =.0014)
- Whether the tutor stated that a body needs to be chosen (20 cases) or not (16 cases). (χ^2=6.1, p=.014)

Because these features are nested, we believe it is really the first feature (errors) that makes a difference. That is, the students who chose a body incorrectly are a subset of the students who chose a body, who are in turn a subset of the students who heard the tutor state that a body needs to be chosen. There are enough gainers in the smallest set to cause all three sets to be reliably associated with gains. This interpretation is consistent with the fact by the time the knot rule comes up, the tutor and student have already discussed many times the need to choose a body, so mentioning it one more time probably doesn't make much difference. Thus, it is more likely that the gains were caused by making errors.

The following features were *not* correlated with gain:

- Whether the tutor stated that the knot is an object.

- Whether the tutor explained why the student's choice of body is wrong (e.g., by demonstrating that choosing a weight as a body leads to a dead-end).
- Whether the tutor explained why knots should be used as bodies (e.g., "You want to relate T3 and W1, and that knot is what you need.").
- Whether the tutor stated the rule in general form.

The non-significance of these features shows that the tutors' explanations are *not* associated with gain. Because the tutors always gave explanations after the student made an error, and errors are associated with gains, it might have been thought that it is the explanations and not the errors that are critical for learning. However, the tutors tended to explain the knot rule even when students did not make an error. Apparently these explanations were not effective, because explanations overall were not associated with gains.

Unlike the deceleration rule, there was no value (or harm) in stating the knot rule in a more general form. However, generalization of the knot rule was not required for solving the test problem because it also involved a knot.

The Compound Body Rule

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 rule is, "A set of objects that move together can be considered a single body." It is taught in the context of a problem where two blocks, one sitting on top of the other, slide down a frictionless inclined plane. It is tested in the context of 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 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, the tutors would often (23 of 25 cases) confirm that the two blocks should be treated as a single body and sometimes (9 cases) would even explain why (i.e., because they have the same acceleration or because they move together).

Of the 6 students who gained, all 6 made a mistake or showed uncertainty, whereas of the 19 students who did not gain, most (16) made the correct body choice without comment. The difference was significant ($\chi^2 = 14.035$, $p = .0002$). No other significant differences were found among the other features that we coded for, namely:

- Whether the tutor explained that the two blocks can be considered a body because they move together.
- Whether the tutor made any other explanations (e.g., there is no need to consider internal forces between blocks).
- Whether the tutor stated the rule in a general form.
- Whether the tutor asked the student to choose the body.
- Whether the student mentioned 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).

These findings suggest that in order to learn the compound body rule, the students needed to make a mistake or show uncertainty. As with the knot rule, it is unlikely that learning was caused solely by the explanation that followed the mistake, because explanations themselves were not associated with gains.

A Kinematics Equation

There are several kinematics (time-rate-distance) equations used in physics, and one of them is $s=v_o t+½\, at^2$, 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 is used in a problem where a block starts at rest and slides down an inclined plane for 2 seconds. It is tested by asking how far an object travels during 10 seconds when starting from rest and accelerating at 5 m/s^2.

The tutorial dialogs for this rule had the following general form. None of the students was able to produce a correct version of the equation. Some produced incorrect equations (13 cases), and some could not produce any equation (9 cases). If the student could not supply 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). If the student produced an incorrect equation, the tutors responded in two ways:

- Sometimes (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 2 of the cases, the tutor justified the target equation by deriving it via calculus.
- Sometimes (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 $s=½at^2$. Additionally, in 3 cases, the tutor derived the equation via calculus.

Two features 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 equation was wrong, then they usually (in 4 of 6 cases) gained, which was a significant difference (χ^2=3.8, p=.053). Second, whenever the tutors derived the target equation via calculus, the students never (out of 7 cases) gained, which was a significant difference (χ^2=5.9, p=.015).

Other features that we coded were not associated with gain:

- Whether the student produced an incorrect equation or no equation.
- Whether the student gave an incorrect answer on the pre-test or gave no answer on the pre-test.
- Whether the student made the same mistake they made on the pre-test.
- 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.

These findings suggest that merely correcting a mistaken equation was not sufficient to remedy the buggy knowledge; the tutor should have explained why the error was wrong. When teaching the target equation, the tutors used different kinds of explanations, but using calculus to derive the target equation apparently only confused the students.

The Rotated Axes Rule

When solving physics problems that have forces arrayed in two dimensions, it is sometimes convenient to use coordinate axes that are tilted from their usual horizontal and vertical orientation. For instance, if a block slides down an inclined plane, then simpler equations are produced by making the x-axis parallel to the plane and the y-axis perpendicular to the plane. During training, the rotated axis rule was used on two inclined-plane problems, although we analyzed only the first one. During testing, students were shown an object with forces drawn on it and asked to draw coordinate axes.

Tutors' explanations were usually based on an overly specific version of the rule, namely, that the axes should be rotated to align with the direction of motion. Although this version of the rule applies to the training problems, it does not apply in the test problem, whose body is stationary. The more general version of the rule is to rotate the axes so that one axis is aligned with the vector one seeks. For most problems with moving objects, the sought quantity is usually a kinematic quantity such as acceleration, which is why the overly specific version of the rule usually works. The tutors discussed the general version of the rule only once, although they could have mentioned it on subsequent problems, which we did not analyze.

The tutoring of the rotated axes rule proceeded as follows. In 12 cases, the tutor suggested rotating the axes before giving the student a chance to draw them. In the remaining 5 cases, the student chose the axes and usually (4 cases) chose non-rotated ones. However, even if the tutor did suggest rotating the axes, 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 did choose correctly, they did not express uncertainty. Regardless of how the axes were chosen, the tutors often (16 or 17 cases) gave a short explanation of the rule (e.g., "It'll save you work if you rotate the axes so x aligns with the acceleration.").

The major feature that predicted gains was whether or not the student made an error by drawing a non-rotated axis. Of the 5 students who made the mistake, 4 were gainers. Of the 12 students who did not make a mistake, only 3 were gainers. (χ^2=4.4, p=.036).

Other features that were not associated with gains include:

- Whether the tutor explained the benefits of rotating the axes.
- Whether the tutor mentioned the overly specific version of the rule.
- Whether force vectors were drawn before or after the coordinate axes.
- Whether a problem that required a single, vertical axis had the axis labeled x or y. Labeling it x could be interpreted as rotating the axes by 90 degrees.

As with the knot and compound body rule, it is errors rather than the explanations that are associated with gain.

Discussion

Our original hypothesis was that effective potential learning events could be discriminated from ineffective ones on the basis of general features, such as

1. Who (student vs. tutor) and what (error vs. question) initiated the event?
2. How many key ideas behind the rule were mentioned?
3. How many words were uttered by participants during the discussion?
4. How many times was the correct conclusion first generated by the student?

Only the last two features turned out to be significantly correlated with learning. Feature 3 suggests that long-winded explanations thwart learning. However, discussions generally continue until the tutor judges that the student has learned the rule or the tutor gives up. Thus, short discussions occurred whenever the tutor found it easy to teach the rule to the student, and in those cases, the student tended to apply the rule on the post-test. Thus, it could be lack of learning that causes long discussions and not vice versa.

Feature 4 has two interpretations. First, it suggests that when tutors let students produce the first correct application of the rule themselves, then the students tend also to construct the rule themselves, and this produces success on the post-test. However, it may also be that whenever a tutor sees that a student is not learning a rule despite the tutor's best efforts, the tutor applies the rule instead of embarrassing the student any further, and in these cases, the student tends not to apply the rule on the post-test.

Although interesting, none of these general features explain much of the variance. This suggested examining rule-specific features.

Five rules were analyzed to see what features of the tutorial dialogs discriminated students who learned the rule from students who failed to learn it. The features associated with learning were different with different rules:

- For the knot rule, the compound body rule, and the rotated axes rule, students who made an error tended to gain. For the compound body rule, students who applied the rule correctly but expressed uncertainty also tended to gain. For all 3 rules, if the tutor applied the rule or the student applied the rule correctly without expressing uncertainty, then the students tended not to gain.
- For the deceleration rule, only students who heard a generalization of the rule tended to gain.
- If a student produced an incorrect kinematics equation, explaining why it was wrong produced gains. Correcting the equation without explaining why it was wrong did not produce gains.
- If the target kinematics equation was explained by using calculus to derive it from the definitions of velocity and acceleration, then students did not gain.

From these findings, several general observations can be made.

First, it seems essential that students become aware that they have a knowledge deficit. In the cases of the deceleration and kinematics rules, all students either made an error or got stuck, and thus realized that their knowledge was either incorrect or incomplete. In the cases of the knot, compound body, and rotated axes rules, only some students made errors or got stuck, and they were the ones that tended to gain. If the tutors either applied the rules themselves or provided such strong hints that the students could easily apply the rule correctly, then the students may not have realized that their knowledge was flawed, which could explain why they tended not to gain.

However, even if the students realized that they had a knowledge defect, they did not necessarily learn from the tutoring. When the explanation involved less familiar background knowledge, such as calculus, the explanation may only have confused the students. Two of the rules, the deceleration rule and the rotated axes rule, required some generalization in order to be used successfully on the post-test. In the case of the deceleration rule, when the tutors mentioned the general rule, the students tended to gain.

Remedying misconceptions that produced incorrect kinematics equations seemed to require that the tutor explain why the misconceptions were wrong. This is surprising, given that Sleeman, Kelley, Martinak, Ward and Moore (1989) found that tutors who explained why algebra "mal-rules" were wrong did no better than tutors who merely said that the rule was wrong without explaining why. Their result is consistent with most of the rules in this study, where gains were not associated with explanations of why errors were wrong. Apparently, a few misconceptions must be "untaught" while others can simply be overridden.

From a practical standpoint, two heuristics for tutoring emerge from these data. First, tutors should let students make mistakes instead of usurping the opportunity by giving strong hints or doing the reasoning themselves. This heuristic must be applied with caution, since we only measured the students' learning and not the change in their motivation or interest, which could be negatively affected by letting them make errors. The second heuristic is that different rules may require different emphases during tutorial explanations. This could be a significant challenge for developers of intelligent tutoring systems, since it necessitates finding out, for each rule, what makes that rule difficult to learn.

Acknowledgements

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

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

Conati, C., & VanLehn, K. (1995). A student modeling technique for problem solving in domains with large solution spaces. In J. Greer (Ed.), *Proceedings of the 1995 AI and Education Conference*. Charlottesville, VA: AACE.

Martin, J., & VanLehn, K. (1995). Student assessment using Bayesian nets. *International Journal of Human-Computer Studies*, 42, 575-591.

Sleeman, D., Kelley, 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. (1996). Conceptual and meta learning during coached problem solving. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), ITS96: *Proceeding of the Third International conference on Intelligent Tutoring Systems*. New York: Springer-Verlag.

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.