Minimally Invasive Tutoring of Complex Physics Problem Solving

Kurt VanLehn[1], Collin Lynch[1], Linwood Taylor[1], Anders Weinstein[1],
Robert Shelby[2], Kay Schulze[2], Don Treacy[2], and Mary Wintersgill[2]

[1] LRDC, University of Pittsburgh, Pittsburgh, PA USA 15260
VanLehn@cs.pitt.edu, lht@lzri.com, {collinL,andersw}@pitt.edu
http://www.pitt.edu/~vanlehn/andes.html
[2] US Naval Academy, Annapolis, MD, USA
{Shelby, Schulze, Treacy, MWinter}@usna.edu

Abstract. Solving complex physics problems requires some kind of knowledge for selecting appropriate applications of physics principles. This knowledge is tacit, in that it is not explicitly taught in textbooks, existing tutoring systems or anywhere else. Experts seem to have acquired it via implicit learning and may not be aware of it. Andes is a coach for physics problem solving that has had good evaluations, but still does not teach complex problem solving as well as we would like. The conventional ITS approach to increasing its effectiveness requires teaching the tacit knowledge explicitly, and yet this would cause Andes to be more invasive. In particular, the textbooks and instructors would have to make space in an already packed curriculum for teaching the tacit knowledge. This paper discusses our attempts to teach the tacit knowledge without making Andes more invasive.

1 Objectives

The Andes project [1-3] began with three objectives [4]. The first was to improve the learning of university physics students. This goal has been accomplished. In large-scale field evaluations at the US Naval Academy over the last three years, students who did their homework with Andes learned significantly more than students who did similar homework on paper. These results are discussed below.

The second objective was to see if Andes could be minimally invasive. In particular, could students adopt Andes for doing their homework while virtually nothing else in the physics course changed? The professors would give the same lectures, use the same textbooks, assign the same homework problems and conduct the same labs and recitations. We think that Andes will have a much wider impact if it can be used with many kinds of teaching, both conventional and reformed. This goal has been difficult to achieve, and our progress is discussed below.

The third objective was to test three help systems. The Conceptual Helper is called when Andes decides that the student is unfamiliar with a specific principle of physics or has a misconception. The Conceptual Helper uses "minilessons" that are adapted to the context of the student's problem solving. Students using Andes with the Con-

S.A. Cerri, G. Gouardères, and F. Paraguaçu (Eds.): ITS 2002, LNCS 2363, pp. 367–376, 2002.

ceptual Helper learned more than students using a version of Andes with ordinary hint sequences [5, 6]. Recent work has replaced the expository minilessons of the Conceptual Helper with natural language dialogs run by Atlas, and has shown that this results in an improvement in student understanding [7].

The Self-Explanation Coach coaches students as they study a solved physics problem (i.e., an example). In order to determine if the student has self-explained an example adequately, it monitors the location and latency of the student's visual attention. If it appears that the student fails to self-explain specific key aspects of the example, the coach guides the student in doing so. Some students who used the SE Coach learned more than students who studied the same examples without the coach [8, 9].

The Procedural Helper answers help requests while students are solving problems [10]. In particular, if the student gets stuck and asks, "What do I do next?" the Procedural Helper will suggest a goal or action. If Andes marks a student entry wrong and the student asks, "What's wrong with that?" the Procedural Helper gives advice based on determining whether the error is the result of either an incorrect inference or a correct inference that does not lead to the goal. The Procedural Helper has been evaluated repeatedly, as described below. Although it is improving, it is still not as effective as we would like.

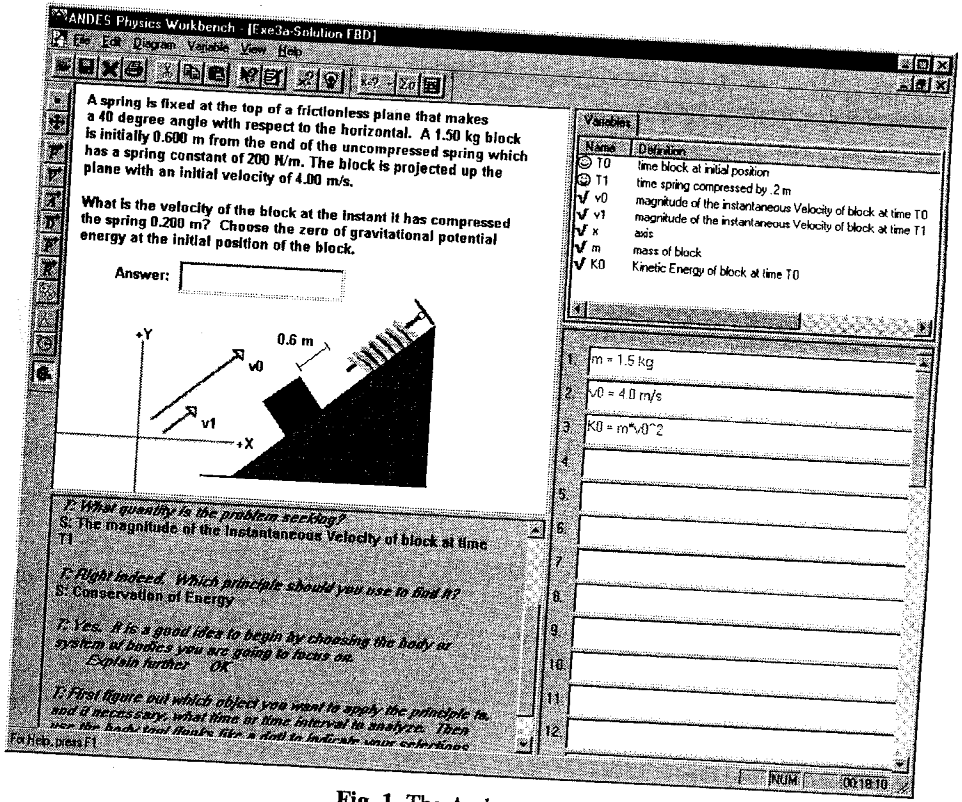
In short, although we have achieved many of our objectives, two remain: reducing the invasiveness of Andes and increasing the effectiveness of the Procedural Helper. This paper reviews our progress towards achieving those goals.

2 The Andes1 User Interface

This paper only discusses the part of Andes that coaches students as they solve problems. A typical physics problem and its solution on the Andes screen are shown in Figure 1. Students read the problem (top of the upper left window), draw vectors and coordinate axes (bottom of the upper left window), define variables (upper right window) and enter equations (lower right window). These are exactly the actions that they should do when solving physics problems with pencil and paper. The main difference between Andes and paper are:

1. Andes gives immediate feedback by turning correct entries green and incorrect entries red. If a red entry is the result of a low-level error (e.g., illegal algebraic syntax), an error message pops up saying so.
2. Andes answers "what should I do next?" and "what's wrong with that?" help requests. Most help requests are answered by canned-text and menu dialogues in the tutor window (lower left). If Andes determines that the student has flawed physics knowledge, it invokes either the Conceptual Helper, which conducts a hypertext-based minilesson, or Atlas, which conducts a natural language dialogue in the tutor window.
3. Andes will solve algebraically the equations that the student has entered, provided that student has entered enough correct ones.
4. Variables must be defined before they are used in equations, and the only way to define certain variables is to draw vectors and/or coordinate axes.

With the exception of the fourth point, which will be discussed later, Andes acts simply as an unobtrusive but helpful piece of paper. This should facilitate transfer to unsupported, paper-based problem solving.



A spring is fixed at the top of a frictionless plane that makes a 40 degree angle with respect to the horizontal. A 1.50 kg block is initially 0.600 m from the end of the uncompressed spring which has a spring constant of 200 N/m. The block is projected up the plane with an initial velocity of 4.00 m/s.

What is the velocity of the block at the instant it has compressed the spring 0.200 m? Choose the zero of gravitational potential energy at the initial position of the block.

Answer:

Name	Definition
☺ T0	time block at initial position
☺ T1	time spring compressed by .2 m
✓ v0	magnitude of the instantaneous Velocity of block at time T0
✓ v1	magnitude of the instantaneous Velocity of block at time T1
✓ x	axis
✓ m	mass of block
✓ K0	Kinetic Energy of block at time T0

1. m = 1.5 kg
2. v0 = 4.0 m/s
3. K0 = m*v0^2
4.
5.
6.
7.
8.
9.
10.
11.
12.

T: What quantity is the problem seeking?
S: The magnitude of the instantaneous Velocity of block at time T1

T: Right indeed. Which principle should you use to find it?
S: Conservation of Energy

T: Yes. It is a good idea to begin by choosing the body or system of bodies you are going to focus on.
Explain further OK

T: First figure out which object you want to apply the principle to and if necessary, what time or time interval to analyze. Then use the body tool [illegible] like a doll to indicate your selections

Fig. 1. The Andes screen.

3 The Andes1 Procedural Help System

Two major versions of Andes have been developed. From the student's view, they differed only in the kind of help they gave when asked "What's next?" or "What's wrong?" That is, they had the same user interface but completely different Procedural Help systems. This section briefly describes the first version of Andes, called Andes1. Subsequent sections discuss its evaluation, then Andes2 and its evaluation.

When the student was stuck and asked Andes1 "What should I do next?", it would first select a target step then give a sequence of hints intended to suggest that action to the student. The hint sequences were similar to those used by many intelligent tutoring systems, but the method of selecting a target step was unique. Andes1 precomputed all possible solutions to the problem and stored them as a Bayesian network called the *solution graph* [11, 12]. The nodes in the solution graph represented goals, facts and other propositions, as well as the inferences (rule applications) that connected the propositions. Andes1 searched the solution graph for a target step as follows [10]. Starting from the student's most recent correct action, it found a goal in the network that was likely to be the one dominating that action. It then searched for another action dominated by that goal that had not yet been entered and was probably one that the student didn't know how to do. This was selected as the target step.

If the student made an incorrect entry and asked Andes1 "What's wrong with that?", it again selected a target step and gave a hint sequence. The target step was selected by comparing the incorrect entry to all possible entries of that type and selecting the closest matching one [13]. The position of the target step in the solution graph was not used.

4 Evaluations of Andes1

The first full-scale evaluation of Andes1 occurred during the fall of 1999 as part of the regular US Naval Academy physics course. For about 6 weeks, 173 students used Andes1 to do their homework, and 162 students did their homework using pencil and paper. The Andes1 students scored significantly higher ($t=2.2$, $p=.036$) on a midterm exam given at the end of the intervention period, and the effect size was a modest 0.21 standard deviations [14].

More detailed evaluations were undertaken to find ways to improve Andes1. Students at the University of Pittsburgh who had recently taken physics were asked to solve problems on Andes while giving a verbal protocol. They were often totally lost and seldom found Andes' help useful.

In order to get a more formal evaluation at this level of detail, we extracted from the log files 40 episodes where a student had asked Andes1 "what's wrong?" with an equation. We printed screen snapshots just before Andes1 gave its advice. Working independently, the three USNA physicists wrote on the snapshots the advice that they would give to the student. Often the advice was somewhat different, but on 21 snapshots their advice was the same. However, Andes' advice was the same as the physicists' advice on only 3 of the 21 snapshots.

Two patterns stood out in the physicists' advice. One was that they often insisted that students do any steps that they had skipped in the procedure for applying a principle. For instance, many students had trouble writing a correct equation for Newton's second law, and they had skipped drawing the force and acceleration vectors. Instead of helping the students correct their equation, the physicists would usually ask the students to draw the missing vectors.

The second pattern was that when the student was lost, the physicists would not just select a target step and hint it as Andes would. Instead, they would help the students infer a target step themselves via a dialogue such as the following:

Tutor:	What quantity is the problem seeking?
Student:	The acceleration of the car.
Tutor:	What principle should you use to find it?
Student:	Newton's second law.

If the student had not yet applied Newton's second law, then the physicists would coach the hypothetical student through the steps in its procedure. If the student had already applied Newton's second law, the physicists would say, "I see you've applied it already as equation 4. What quantities in the equation are unknown?" When the student answers, "the frictional force on the car," the process repeats and the physicists ask for a principle to find that quantity, etc.

Lastly, we discovered that Andes would frequently give outlandish advice when asked "What's wrong?" with an incorrect equation. Often it would suggest replacing a piece of the equation with a specific number or expression, creating an equation that none of the physicists could recognize. This turned out to be due to its basic algorithm for selecting a target step for "What's wrong" help. Even if the student was just beginning to solve the problem, it would consider target equations from the very end of the solution. Moreover, it would consider all possible algebraic combinations of correct equations even if that combination didn't participate in the solution. Consequently, it would hint writing an equation that the student wasn't intending to write and should never write. Unfortunately, when students took Andes' advice and entered the suggested equation, it would turn green (correct). Students probably found this terribly confusing.

The second full-scale evaluation occurred during fall of 2000 at the US Naval Academy. The major goal was to increase the amount of physics covered, so the intervention lasted longer (10 weeks), covered more problems (60) and covered more topics. We also made some limited changes to cure the problems found in the analyses of the 1999 log files: (1) Students were required to draw vectors rather than define vector variables without drawing them. (2) The plan inference method described in [10] was replaced with a simpler technique. These changes left Andes1 giving essentially the same feedback and help as the 1999 version. Nonetheless, the Andes1 students scored significantly higher on a midterm exam just after the intervention ($t=7.74$, $p<.00001$), and this time the effect size was a satisfying 0.92 standard deviations [14]. However, Andes' advice still differed substantially from the advice given by the physicists, so it seemed that there was still room for improvement.

5 The Objectives of Andes2

The physicists clearly had a goal hierarchy in mind when they gave advice to the students. The lower levels of the goal hierarchy corresponded to well-known methods for applying principles, which are printed in textbooks. For instance, the method for applying Newton's second law to an object is to draw all the forces acting on the object, draw its acceleration, draw coordinate axes and write an equation. The upper levels of the goal hierarchy seemed to correspond to a search that starts at the sought quantity, applies a principle containing the sought, and then recurses to find values for any unknowns in the principle's equation. This corresponds to a backwards search strategy used by many expert physics problem solving systems [e.g., 15]. Although empirical studies have not precisely identified the strategies used by human experts, it appears that they use this backwards search some of the time, but more often hold principle applications in memory until they have planned out a complete solution, then they write equations as they solve them [16]. Our physicists' advice was consistent with just one strategy, backwards search, so we decided to teach that, along with the well-known procedures for applying principles.

Our first task was to invent a display that reified the goal hierarchies. When we implemented such a display (two of them, in fact), the physicists carefully considered them but ultimately rejected them as too invasive. For students to use these displays, they would have to be explicitly taught the backwards search algorithm, a vocabulary for specifying goals and some tools for navigating and editing the goal hierarchies. This seemed to the physicists to require too much class time. Even if they changed their curriculum, they doubted that any of their colleagues would be willing to. In short, the goal of minimal invasiveness was incompatible with the standard technique of reifying goals and explicitly teaching a problem solving strategy.

Thus, we decided to have Andes teach the problem solving strategy in the same way that the physicists seemed to teach it: as part of their advice when asked for help. That is, when a student asks "What should I do next?" Andes should engage the student in the same question-answer dialogue that the physicists did.

6 The Implementation of Andes2

Many changes were required in order to implement Andes2. First, we had to restructure the physics knowledge base. Instead of a flat set of several hundred inference rules as used in Andes1, the new knowledge base was organized as about 100 principles, each with its textbook method for applying it.

The solution graph was restructured as a *bubble graph* and a set of *method graphs*. The bubble graph is composed of two kinds of nodes: nodes representing quantities and nodes representing principle applications. A node representing a principle application is linked to the nodes representing the quantities that appear in the principle application's equation. Associated with each principle application node is a graph of propositions similar to the ones used in Andes1. This graph, called the method graph, represents how to actually accomplish the application of the principle.

When the student asks "What should I do next?", Andes2 conducts a dialogue based on the solution graph. It always starts by asking the student, "What quantity does the problem seek?" It offers the student a hierarchical menu of all physics quantities. The student gets negative feedback and another chance if the student fails to pick a quantity that is actually sought by the problem. Otherwise, the tutor asks, "What principle should be used to find it?" and offers a hierarchical menu of all principle applications. To evaluate the student's selection, it starts at the sought quantity's node in the bubble graph and sees if the student's selected principle application is indeed on a solution path for this problem. If not, Andes2 says so and asks the student to try again. If the selected principle application is appropriate and not yet finished, then Andes2 enters the method graph for that principle. It locates the first unaccomplished step, and composes a hint sequence for that step. These hints drive the rest of the dialogue. On the other hand, if the principle has already been applied, Andes2 says so and asks the student which unknowns in the equation should be determined next. The student selects a quantity, and Andes2 uses the bubble graph to check that it is indeed an unknown in the principle application. From this point on, the process recurses. Eventually the tutor and the student end up at an unfinished principle application, traverse its method graph, select a target step and accomplish it.

Since Andes2 locates a target step by asking a sequence of questions of the student, it has no need to guess a target step based on probabilistic reasoning. Thus, Andes2 does not use a Bayesian network. We would revive it if Andes needed estimates of the student's mastery of principles in order to decide which problem to assign next and whether to go on to the next chapter. However, having different students do different homework was believed to be too invasive for the US Naval Academy. For instance, students who were assigned more physics problems could justifiably complain that Andes hurt their grades in their other courses.

We also revised the help that Andes gives when students ask, "What's wrong with that?" Instead of finding a closest matching target step then hinting the difference between it and the student's step, Andes2 has a set of error handlers. Each can recognize a specific type of error. For instance, there is an error handler for using the wrong time specifier on a given quantity, which is usually due to a misreading of the problem statement. There is an error handler for failing to include a minus sign on a vector component when the vector is parallel to the axis; this is usually due to a misunderstanding of vector algebra. Each error handler has hints and/or minilessons designed to remedy the error and any misconceptions that might underlie it. If none of the error handlers recognizes the student's incorrect entry, then the student is advised to ask, "What should I do next?"

Joel Shapiro, a physicist who visited our group for a year, implemented a fundamental change in the way Andes recognizes equations [17]. Andes1 precomputed a table of all algebraic combinations of principle applications. This allowed it to recognize T_y-mg=m*a_y even though it is composed of several principle applications: T_y+W_y=m*a_y and W_y = -W = -m*g. This precomputation became intractable as Andes1 began to handle more complex problems. Andes2 recognized which primitive equations have been combined to form the student's equation by taking the gradient of the student's equation at a particular point, then seeing whether there is a set of principle applications such that the gradients of their equations sum to the gradient of the student's equation. This made it possible to handle problems with several hundred principle applications.

7 Evaluations of Andes2

Andes2 was evaluated in fall 2001 at the US Naval Academy. The intervention lasted around 12 weeks. Once again, the Andes students learned significantly more than the control students (t=3.14, p=.0012). The 0.52 effect size was respectable but less than Andes1 achieved in Fall 2000.

It was quite clear, both from log file analysis and from the comments of the students to the instructors, that Andes2 was simply young software. Although we had succeeded in removing bugs that would crash the system or cause obvious malfunctions, many pedagogical bugs remained. For instance, since all the hint sequences were new, many of the hints were phrased in confusing ways. The error handlers sometimes misrecognized errors. There were hundreds of these little "pedagogical bugs." Pilot testing would have uncovered them, but we could only run a few pilot subjects before the Naval Academy semester began. In contrast, Andes1 was tuned using log files from several hundred subjects by the time it reached the Fall 2000 evaluation, so it was much more mature software.

The instructors were happy to report that Andes2 did not suggest outlandish equations when students asked what was wrong with one of their equations. Moreover, student acceptance of the tutor appears not to have been hurt, and may even have gone up slightly. When asked at the end of the intervention if they would chose to continue using Andes if they could, 33% of the Fall 2001 students reported that they would versus 28% of the Fall 2000 students. Although this is good news, there is still room for improvement.

8 Conclusions and Future Work

We are finally beginning to understand the problem faced by Andes. Physics knowledge can be divided into (1) principles and the multi-step methods used to applying them, and (2) some kind of problem solving strategy, such as backwards search, that is used to *select* a principle application. All the stakeholders agree that physics students should learn principles and the methods for applying them. Indeed, recent textbooks print the application method for Newton's law, the application method for translational kinematics, and the application methods for many other major principles. However, there is no consensus on whether to explicitly teach a strategy for *selecting* principle applications. On the one hand, many successful tutoring systems have been built around the common wisdom that one should first find out what tacit knowledge is required for successful problem solving, then design a tutoring system that teaches that knowledge explicitly [e.g., 18]. On the other hand, today's experts probably acquired their principle selection strategy via implicit learning. There are even computational models of how such implicit learning could occur [19, 20]. Perhaps it would be best if students were not explicitly taught a principle selection strategy, but instead learned it implicitly. Moreover, explicit teaching of a principle selection strategy would be more invasive than implicit teaching. It would require augmentation of the textbooks, changes to the lectures, mastery of a notation for goals and reallocation of precious student time from learning principles to learning strategies.

Since we do not know whether implicit or explicit learning is better for principle selection knowledge, we designed Andes2 as a compromise between them. It offers explicit teaching of the backwards search strategy for selecting principles, but only when asked.

Clearly, the next major goal for the Andes project is an experiment that measures the effectiveness of different amounts of explicit teaching of principle selection knowledge. The current plan is to use 3 versions of Andes2. One has help turned off, so that students must learn principle selection knowledge implicitly. The second is the current version, which teaches some but not all the necessary principle selection knowledge. The third is a new version, which explicitly teaches principle selection knowledge. It will reify goals and be accompanied by printed materials with copious examples. A modified curriculum will be developed that inserts explicit teaching of principle selection knowledge at key points (e.g., when students are able to solve single-principle problems easily, but have not yet begun to solve multiple-principle problems). All three versions of Andes2, along with the printed instructional material, should be thoroughly pilot tested in order to remove pesky pedagogical bugs like the ones that plagued the Fall 2001 version.

Depending on the results of this experiment, we may move to the next obvious study, which is seeing which versions of the tutor are acceptable to instructors and students. That is, how much better do the more explicit tutors have to be in order to justify their invasiveness?

At issue here is perhaps one of the oldest pieces of advice in the ITS literature: explicate tacit knowledge, then teach it. We hope to discover whether that advice is correct for physics.

Acknowledgements. This research was supported by grant N00014-96-1-0260 from ONR Cognitive Sciences.

References

1. Gertner, A.S. and K. VanLehn, Andes: A coached problem solving environment for physics, in *Intelligent Tutoring Systems: 5th international Conference, ITS 2000*, G. Gautheier, C. Frasson, and K. VanLehn, Editors. 2000, Springer: New York. p. 133-142.
2. Schulze, K.G., et al., Andes: An intelligent tutor for classical physics. *The Journal of Electronic Publishing*, 2000. 6(1).
3. Schulze, K.G., et al., Andes: An active learning intelligent tutoring system for Newtonian physics, in *THEMES in Education*. 2000, Leader Books: Athens, Greece. p. 115-136.
4. VanLehn, K., Conceptual and meta learning during coached problem solving, in *ITS96: Proceeding of the Third International conference on Intelligent Tutoring Systems.*, C. Frasson, G. Gauthier, and A. Lesgold, Editors. 1996, Springer-Verlag: New York.
5. Albacete, P.L. and K. VanLehn, The Conceptual Helper: An intelligent tutoring system for teaching fundamental physics concepts, in *Intelligent Tutoring Systems: 5th International Conference, ITS 2000*, G. Gauthier, C. Frasson, and K. VanLehn, Editors. 2000, Springer: Berlin. p. 564-573.
6. Albacete, P.L. and K. VanLehn, Evaluation the effectiveness of a cognitive tutor for fundamental physics concepts, in *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, L.R. Gleitman and A.K. Joshi, Editors. 2000, Erlbaum: Mahwah, NJ. p. 25-30.
7. Rose, C.P., et al., Interactive conceptual tutoring in Atlas-Andes, in *Artificial Intelligence in Education: AI-Ed in the Wired and Wireless future*, J.D. Moore, C. Redfield, and W.L. Johnson, Editors. 2001, IOS: Washington, DC. p. 256-266.
8. Conati, C. and K. VanLehn, Further results from the evaluation of an intelligent computer tutor to coach self-explanation, in *Intelligent Tutoring Systems: 5th International Confer-*

ence, ITS 2000, G. Gauthier, C. Frasson, and K. VanLehn, Editors. 2000, Springer: Berlin. p. 304-313.

9. Conati, C. and K. VanLehn, Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, 2000. 11: p. 398-415.
10. Gertner, A., C. Conati, and K. VanLehn, Procedural help in Andes: Generating hints using a Bayesian network student model., in *Proceedings of the 15th national Conference on Artificial Intelligence*. 1998.
11. Conati, C., A. Gertner, and K. Vanlehn, Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interactions*, in press.
12. Conati, C., et al., On-line student modeling for coached problem solving using Bayesian networks, in *User Modeling: Proceedings of the Sixth International conference, UM97*, A. Jameson, C. Paris, and C. Tasso, Editors. 1997, Spring Wien: New York.
13. Gertner, A.S., Providing feedback to equation entries in an intelligent tutoring system for Physics, in *Intelligent Tutoring Systems: 4th International Conference*, B.P. Goettl, et al., Editors. 1998, Springer: New York. p. 254-263.
14. Shelby, R.N., et al. The Andes Intelligent Tutor: an Evaluation. In *Physics Education Research Conference*. 2001. Rochester, NY.
15. Bundy, A., et al., Solving mechanics problems using meta-level inference, in *Proceedings of the Sixth International Joint Conference on AI*. 1979, Morgan Kaufmann: San Mateo, CA. p. 1017-1027.
16. Priest, A.G. and R.O. Lindsay, New light on novice-expert differences in physics problem solving. *British Journal of Psychology*, 1992. 83: p. 389-405.
17. Shapiro, J.A., Algebra subsystem for an intelligent tutoring system. *International Journal of Artificial Intelligence in Education*, submitted.
18. Anderson, J.R., et al., Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences*, 1995. 4(2): p. 167-207.
19. VanLehn, K., R.M. Jones, and M.T.H. Chi, A model of the self-explanation effect. *The Journal of the Learning Sciences*, 1992. 2(1): p. 1-59.
20. Elio, R. and P.B. Scharf, Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization. *Cognitive Science*, 1990. 14: p. 579-639.