2007

Can Help Seeking be Tutored? Searching for the Secret Sauce of Metacognitive Tutoring

Ido Roll
Carnegie Mellon University

Vincent Aleven
Carnegie Mellon University

Bruce M. McLaren
Carnegie Mellon University

Kenneth R. Koedinger
Carnegie Mellon University

Roll, I., Aleven, V., McLaren, B. M. & Koedinger, K. R. (2007). Can Help Seeking Be Tutored? Searching for the Secret Sauce of Metacognitive Tutoring. In the *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (AIED 2007), 203-210.

Can Help Seeking Be Tutored? Searching for the Secret Sauce of Metacognitive Tutoring

Ido ROLL, Vincent ALEVEN, Bruce M. MCLAREN, Kenneth R. KOEDINGER
Human Computer Interaction Institute, Carnegie Mellon University

Abstract. In our on-going endeavor to teach students better help-seeking skills we designed a three-pronged Help-Seeking Support Environment that includes (a) classroom instruction (b) a Self-Assessment Tutor, to help students evaluate their own need for help, and (c) an updated version of the Help Tutor, which provides feedback with respect to students' help-seeking behavior, as they solve problems with the help of an ITS. In doing so, we attempt to offer a comprehensive help-seeking suite to support the *knowledge*, *skills*, and *dispositions* students need in order to become more effective help seekers. In a classroom evaluation, we found that the Help-Seeking Support Environment was successful in improving students' declarative help-seeking knowledge, but did not improve students' learning at the domain level or their help-seeking behavior in a paper-and-pencil environment. We raise a number of hypotheses in an attempt to explain these results. We question the current focus of metacognitive tutoring, and suggest ways to reexamine the role of help facilities and of metacognitive tutoring within ITSs.

Keywords. Help Seeking; Self-Assessment; Metacognition; Self-Regulated Learning; Intelligent Tutoring Systems; Cognitive Tutors; Empirical Study

Introduction

One of the challenges students face while working with an Intelligent Tutoring System (ITS), as part of regulating their learning, is choosing what actions to perform: using an online help resource, approaching a teacher or peer, or attempting to solve the next problem step. Seeking the right form of help at the right time is known to be associated with better learning [1; 17]. However, students' help-seeking behavior is far from ideal [3]. Within an ITS, students often ask for over-elaborated help when none or little is needed, but avoid asking for necessary help. This maladaptive help-seeking behavior appears to be consistent across domains and students [12].

One approach to help students improve their help-seeking behavior is to design the system so that it limits their opportunities for maladaptive help seeking. For example, Wood [17] created a "contingent tutor" that adapts the hint level to the student's proficiency, and the Cognitive Tutor [8] has a built-in two seconds delay that prevents repeated fast hint requests. However, this approach does not necessarily lead to an improvement in students' help-seeking skills. A different approach separates the cognitive and the metacognitive demands of the task. For example, Reif and Scott [10] and Gama [7] separate the practice opportunities of monitoring or planning skills from those of domain skills. However, it is important that the help-seeking behavior is practiced within the learning context.

In this project we attempt to help students acquire and spontaneously apply better help-seeking skills that persist beyond the scope of the tutoring system itself. We do so in the context of the Geometry Cognitive Tutor, which, like other Cognitive Tutors [8], gives students on time, tailored feedback on their problem-solving actions. It does so by tracing students' actions relative to a cognitive model. The Geometry Cognitive Tutor has two help mechanisms: (i) several levels of contextual hints are available for students, with the most elaborated one giving away the answer, and (ii) the glossary is a searchable geometry knowledge base, much like an online dictionary.

Following the Cognitive Tutor pedagogy [8], our first attempt was to teach help seeking by giving students feedback on their help-seeking behavior (specifically, their help-seeking errors). The Help Tutor [11], a Cognitive Tutor in its own right, identifies recommended types of actions by tracing students' interaction with the Geometry Cognitive Tutor in real time relative to a metacognitive help-seeking model [1]. When students perform actions that deviate from the recommended ones, the Help Tutor presents a message that stresses the recommended action to be taken. Messages from the metacognitive Help Tutor and the domain-level Cognitive Tutor are coordinated, so that the student receives only the most helpful message at each point [2].

Previous evaluations of the Help Tutor led to mixed results. On the one hand, the Help Tutor did as planned – gave feedback on learning events that were associated with poorer learning outcomes, and led to a reduction in help-seeking errors. On the other hand, working with the Help Tutor did not lead to improved learning of domain knowledge, nor to better declarative knowledge of the ideal help-seeking behavior [11].

1. The Help-Seeking Support Environment

These results suggest that a more comprehensive approach is required in order to help students become better help seekers, one that includes substantial declarative and dispositional components, in addition to a procedural one. To address these needs we designed the Help-Seeking Support Environment (HSSE), which includes the following components:

The Updated Help Tutor stresses help-seeking principles and benefits, and not merely the error and the recommended action. For example, the message ""Slow down, slow down. No need to rush" was changed to: "It may not seem like a big deal, but hurrying through these steps may lead to later errors. Try to slow down."

The Self-Assessment Tutor. As shown by Tobias and Everson [15], the ability to correctly self-assess one's own ability is correlated with strategic use of help. While

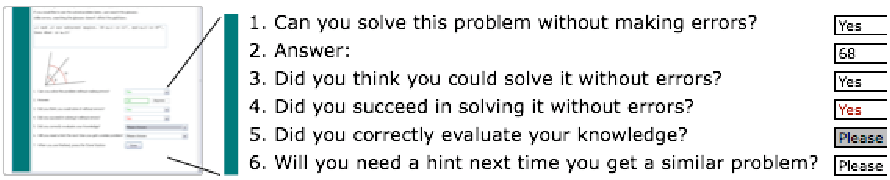

Figure 1. The Self Assessment tutor scaffolds self-assessment in four stages: *Prediction*, in which the student predicts whether she knows how to solve the problem (Q. 1); *Attempt*, in which the student attempts to solve the problem (Q. 2); *reflection*, in which the student contrasts her actual performance with her prediction (Q. 3-5); and *projection*, in which the student projects from the existing experience on future need for help when encountering similar problems (Q. 6).

working with the Self-Assessment Tutor, students are asked to assess their ability on the target set of skills, compare this assessment to their actual performance, and make appropriate choices regarding their subsequent learning process ([13], see Figure 1).

Declarative instruction. As White and Frederiksen demonstrated [16], reflecting in the classroom environment on the desired metacognitive process helps students internalize it. With that goal in mind, we created a short classroom lesson about help seeking with the following objectives: to give students a better declarative understanding of desired and effective help-seeking behavior (e.g., “take the time to think before you act”); to improve their dispositions and attitudes towards seeking help (e.g., “Struggling is part of the learning process. You will not learn by guessing or abusing hints, even if you get the answer right”); and to frame the help-seeking knowledge as an important learning goal, alongside knowledge of geometry. The instruction comprises a 4 minutes video presentation with examples of productive and faulty help-seeking behavior and the main help-seeking principles, followed by 5-10 minutes of teacher-led discussion.

The HSSE curriculum begins with the declarative instruction, followed by interleaved Self Assessment and Cognitive Tutor + Help Tutor sessions, with the Self Assessment sessions taking about 10% of the students’ time (Table 1). In order to help students notice the broad relevance and domain-independent nature of the help-seeking knowledge, we made the HSSE available across two instructional units.

Table 1: The study procedure. ⓘ - Declarative instruction; ✎ - Self-assessment preparatory session. Table is not to scale, i.e., self-assessment and instructional activities took about 10% of the class time

Week:		1	2	3	4		5-9	10	11	12	13	
		Unit 1 (Angles)						Unit 2 (Quadrilaterals)				
Help group	Pretest 1	ⓘ	✎	✎	✎	Posttest 1, Pretest 2	Break	ⓘ	✎	✎	✎	Posttest 2
		Cognitive Tutor + HSSE						Cognitive Tutor + HSSE				
Control group		Cognitive Tutor						Cognitive Tutor				

2. Method

Participants. The HSSE was evaluated with 67 students from 4 classrooms instructed by 2 teachers, from a rural vocational school. All students, 10th and 11th graders, were enrolled in the Cognitive Tutor Geometry class, and thus were familiar with the Cognitive Tutor and its interface.

Design. Since the HSSE includes teacher-led classroom instruction that is not given to the Control condition, the study was done in a between-class fashion. Two classes (with 29 students) were assigned to the Help condition (Cognitive Tutor + HSSE) with the remaining two classes (38 students) were assigned to the Control condition.[1] Classes were assigned in consultation with the teachers, attempting to

[1] An additional class was assigned to the Help condition. We discarded its data due to lack of effort according to the teacher (who reported that the students did not make a serious effort to do well on the tests) and the data. For example, students in this class left blank as many as 38% of the problems on posttest1, compared with only 11% in the other classes. This is probably due to lack of effort, since most of the items left blank were concentrated at the end of the form, regardless of the difficulty level of the counter-balanced problems. Including the class in the analysis does not affect the results qualitatively.

control for number of students and time of day. One class that was reported by the teachers to be of lower ability was assigned to the Help condition. Each teacher taught classes in both conditions.

Procedure. The study spanned a period of three months and two instructional units, during which students in the Help condition used the HSSE for about 15 academic hours (see table 1). Each instructional unit took about a month, with a month in between during which students prepared for the standardized state test. The declarative instruction on help seeking was given twice during the study period in the Help condition classes, at the beginning of each unit. The total time spent on the study was kept constant across classes and conditions.

Assessment design. Students' geometry knowledge was assessed three times across the study: Prior to unit 1 (pretest on unit 1), in between the two units (posttest on unit 1 and a partial pretest on unit 2), and following unit 2 (posttest on unit 2). The tests included "regular" geometry problem-solving items as well as a number of deep-understanding measures: (i) *reason items*. Students were asked to state the theorem they used (ii) *data insufficiency items*. Students were told (on all problems) that if there is not enough information to answer the problem they should state so by writing 'No'. About 10% of all steps in the tests lacked sufficient information. (iii) *conceptual understanding items*. Unit 2 included items on which students needed to demonstrate conceptual understanding, for example, by matching up diagrams and geometric properties with given geometric shapes.

Students' help-seeking behavior was evaluated using embedded hints in the test [11]. Several test items appeared in three hint conditions, counterbalanced between forms: conventional *No hint* items; *Request hint* items, which contained a hint that was covered by a sticker (students were told that removing the sticker would cost them 10% of the item's score); and *Free* and open hint items (see Figure 2).

Last, students' declarative help-seeking knowledge was evaluated using hypothetical help-seeking dilemmas. These multiple-choice questions asked students to choose the appropriate action to perform in response to situations such as repeated errors, easy steps, verbose hints, etc. These situations were not discussed during the instruction.

The tests were piloted for difficulty level and comprehensibility with students from a parallel class at the same school.

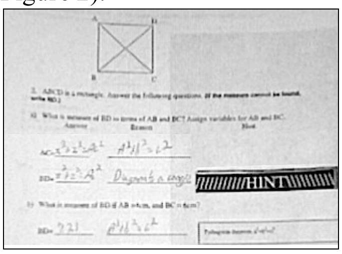

Figure 2. A typical test item with three hint conditions: No hint, Request hint (students were told that removing the sticker will lead to a 10% reduction of their grade on that step), and Free hint. Hint conditions were counterbalanced between test forms.

3. Results

In the study we addressed the following questions:
(a) Did the HSSE affect learning at the domain level? (b) Did it affect help-seeking behavior as measured by the use of embedded hints in the paper test? And (c) did it affect declarative knowledge of good help-seeking? Due to absences, 53, 52 and 43 forms (of the 67 registered students) were collected for the three respective tests, divided about evenly between conditions.

Learning gains. As detailed above, the test included common geometry problem-solving items and robust-learning items in the form of Reason, Data Insufficiency items, and in unit 2, also Conceptual Understanding items. As seen in Figure 3, both groups improved significantly from pretest (time 0) to posttest1 (time 1) on the Problem solving and Reason items (Problem solving: $t(45)=3.1$, $p=0.004$; Reason: $t(45)=5.6$, $p<0.0005$). The scores at posttest2 (time 2) cannot be compared to those at posttest1, since they pertain to a different instructional unit. The only item types at time 2 to have a pretest at time 1 were the Conceptual items, in which significant learning is observed ($t(34)=6.8$, $p<0.0005$). No learning was observed on Data Insufficiency items.

The HSSE had no effect on any of the scores. There were no differences in learning between the groups, with the Help group scoring a bit lower on the procedural problem-solving items on all tests, as was expected, given that one of the classes in this condition included many lower-achieving students. There was also no significant interaction that includes condition.

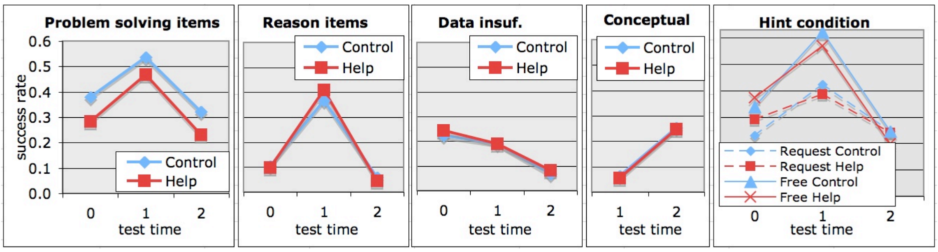

Figure 3. Learning gains on the different measures. From left to right: Problem solving items, Reason items, Data Insufficiency items, Conceptual items, and items with embedded hints. Posttest 2 evaluated a different unit form posttest 1, and thus used a different test. While there is significant learning from pre to post, there are virtually no differences between conditions that are not accounted for by pretest differences.

Hint behavior. A small number of test items had a hint manipulation, counter-balanced between forms: No hint, Request, and Free (see Figure 2.) To control for domain-level difficulty, we measured the added value of having a hint by computing the ratio between scores on items with a hint (either Request or Free) and items that did not have a hint. This measure had been validated earlier by showing that it correlates with online help-seeking behavior in the tutor [11]. As before, no differences were observed between conditions. Also, no interaction that includes condition reached significance.

Figure 3 shows scores on items with hints. Overall, 12 measures of hint scores were used (2 hint types * 2 conditions * 3 test-times). Scores on items with hints were lower than on the same items with no hints on 7 out of the 12 hint measures. In other words, interestingly, having some form of a hint hindered performance on more than half of the measures in the paper test. This is probably due to the novelty of having hints in the tests, although this was not the case in previous uses of this method.

Figure 4. Significant differences in help seeking declarative assessment show that the Help group gained a better understanding of the help-seeking process.

Declarative Knowledge of Help Seeking. The only significant difference between the groups was found in the Declarative Help-Seeking knowledge assessment, evaluated by means of the hypothetical help-seeking dilemmas questionnaire. The Help group students scored 74% whereas the Control group students scored only 47% on that test (Figure 4; $F(1,31)=6.5$, $p<0.02$)[2]. Thus, students in the Help

[2] Not all students filled in the questionnaire. When including all students in the analysis, including those who skipped it, the effect still holds, although is no longer significant (Help: 55%; Control: 37%; p=0.14).

condition demonstrated better understanding of the help-seeking process and not merely greater inclination to use hints.

4. Conclusions

This paper discusses a comprehensive metacognitive environment to support help-seeking behavior, and the outcomes of a classroom evaluation done with the environment. The environment supports and links three types of help-seeking knowledge: declarative, procedural, and dispositional. It combines three instructional approaches: classroom discussion, preparatory self-assessment sessions, and feedback on help-seeking errors in the main tutoring environment (the Geometry Cognitive Tutor). A classroom evaluation with 67 students revealed that use of the HSSE contributed only to students' declarative help-seeking knowledge. We found no differences between the groups with respect to their domain-specific learning or their help-seeking behavior on the paper-and-pencil test.

These somewhat disappointing results raise an important question: Why did the environment not lead to an improvement in learning and in help-seeking behavior in the paper-test measures? One possible explanation may be that the HSSE imposes excessive cognitive load during problem solving. Clearly, the learning process with the HSSE is more demanding compared to that with the conventional Cognitive Tutor alone, since more needs to be learned. However, much of the extra content is introduced during the classroom discussion and self-assessment sessions. The only extra content presented during the problem-solving sessions are the Help Tutor's error messages, but they are not expected to increase the load much, especially given that a prioritization algorithm makes sure students receive only one message at a time (either from the Help Tutor or the Cognitive Tutor).

The HSSE is likely to have improved students' help-seeking behavior while working with it. Although we have not yet analyzed the log files of this study, in a previous study, the Help Tutor by itself (without the additional components of the HSSE) was shown to have an effect on online behavior [11]. Also, the Help Tutor makes it harder to commit certain types of help-seeking errors (such as immediate repeated hint requests) due to its feedback. It is therefore reasonable to assume that students in the Help group committed fewer help-seeking errors than Control group students. But if that is so, why then did the improved online help-seeking behavior not lead to improved learning gains? Several hypotheses can be put forward in this regard, forming two lines of reasoning: *The role of help seeking in ITS*, and *the focus of metacognitive tutoring in ITS*.

The role of help seeking in ITS. Hints in tutoring systems have two objectives: to promote learning of challenging skills, and to help students move forward within the curriculum (i.e., to prevent them from getting stuck). While the latter is achieved easily with both the Cognitive Tutor and the HSSE, achieving the first is much harder. Surprising as it may sound, it is not yet clear what makes a good hint, and how to sequence hints in an effective way. It is possible that the hints as implemented in the units of the Cognitive Tutor we used are not optimal. For example, there may be too many levels of hints, with each level adding too little information to the previous one. Also, perhaps the detailed explanations are too demanding with regard to students' reading comprehension ability. It is quite possible that these hints, regardless of how they are being used, do not contribute much to learning. Support for that idea comes from Schworm and Renkl [14], who found that explanations offered by the system impaired learning when self-explanation was required. The Geometry Cognitive Tutor prompts for self-explanation in certain units. Perhaps elaborated hints are redundant, or even damaging, when self-explanation is required.

It is possible also that metacognitive behavior that we currently view as faulty may actually be useful and desirable, in specific contexts for specific students. For example, perhaps a student who does not know the material should be allowed to view the bottom-out hint *immediately*, in order to turn the problem into a solved example. Support for that idea can be found in a study by Yudelson et al. [18], in which medical students in a leading med school successfully learned by repeatedly asking for more elaborated hints. Such "clicking-though hints" behavior would be considered faulty by the HSSE. However, Yudelson's population of students is known to have good metacognitive skills (without them it is unlikely they would have reached their current position). Further evidence can be found in Baker et al. [4], who showed that some students who "game the system" (i.e., click through hints or guess repeatedly) learn just as much as students who do not game. It may be the case that certain gaming behaviors are adaptive, and not irrational. Students who use these strategies will insist on viewing the bottom-out hint and will ignore all intermediate hints, whether domain-level or metacognitive. Once intermediate hints are ignored, better help-seeking behavior according to the HSSE should have no effect whatsoever on domain knowledge, as indeed was seen. It is possible that we are overestimating students' ability to learn from hints. Our first recommendation is to re-evaluate the role of hints in ITS using complementary methodologies such as log-file analysis (e.g., dynamic Bayes nets [6]); tracing individual students, experiments evaluating the effect of different types of hints (for example, proactive vs. on demand), and analysis of human tutors who aid students while working with ITS.

The focus of metacognitive tutoring in ITS. Students' tendency to skip hints suggests that perhaps the main issue is not lack of knowledge, but lack of motivation. For students who ignore intermediate hints, metacognitive messages offer little incentive. While the HSSE can increase the probability that a proper hint level appears on the screen, it has no influence on whether it is being read. Students may ignore the messages for several reasons. For example, they may habitually click through hints, and may resent the changes that the HSSE imposes. This idea is consistent with the teachers' observation that the students were not fond of the HSSE error messages. The test data discussed above provides support for this idea. On 7 out of the 12 hint evaluations students scored lower on items with hints than on items with no hints. A cognitive load explanation does not account for this difference, since the Request hints did not add much load. A more likely explanation is that students chose to skip the hints since they were new to them in the given context. Baker [5] reviewed several reasons for why students game the system. While no clear answer was given, the question is applicable here as well. Pintrich [9] suggests that while appropriate motivation facilitates the use of existing metacognitive skills, other motivations may hinder such productive behavior.

Motivational issues bring us to our final hypothesis. Time Preference Discount is a term coined in economics that describes behavior in which people would rather have a smaller immediate reward over a distant greater reward. In the tutoring environment, comparing the benefit of immediate correct answer with the delayed benefit (if any) of acting in a metacognitively correct manner may often lead the student to choose the first. If that is indeed the case, then students may already have the right metacognitive skills in place. The question we should be asking ourselves is not only how to get students to learn the desired metacognitive skills – but mainly, how to get students to use them.

Acknowledgements. We thank Ryan deBaker, Jo Bodnar, Ido Jamar, Terri Murphy, Sabine Lynn, Dale Walters, Kris Hobaugh and Christy McGuire for their help. This work was supported by the Pittsburgh Science of Learning Center which is funded by the National Science Foundation (#SBE-0354420), by the Graduate Training Grant awarded by the Department of Education (#R305B040063), and by NSF (#IIS-0308200).

References

[1] Aleven, V., McLaren, B.M., Roll, I., & Koedinger, K.R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education* a(16), 101-30

[2] Aleven, V., Roll, I., McLaren, B.M., Ryu, E.J., & Koedinger, K.R. (2005) An architecture to combine meta-cognitive and cognitive tutoring: Pilot testing the Help Tutor. *in proceedings of 12Th International Conference on Artificial Intelligence in Education.* The Netherlands: IOS press.

[3] Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R.M. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research* 73(2), 277-320

[4] Baker, R.S., Corbett, A.T., & Koedinger, K.R. (2004) Detecting Student Misuse of Intelligent Tutoring Systems. *in proceedings of 7Th International Conference on Intelligent Tutoring Systems,* 531-40.

[5] Baker, R.S., Roll, I., Corbett, A.T., & Koedinger, K.R. (2005) Do Performance Goals Lead Students to Game the System? *in proceedings of 12Th International Conference on Artificial Intelligence in Education,* 57-64. Amsterdam, The Netherlands: IOS Press.

[6] Chang, K.K., Beck, J.E., Mostow, J., & Corbett, A. (2006) Does Help Help? A Bayes Net Approach to Modeling Tutor Interventions. *in proceedings of Workshop on Educational Data Mining at AAAI 2006,* 41-6. Menlo Park, California: AAAI.

[7] Gama, C. (2004) Metacognition in Interactive Learning Environments: The Reflection Assistant Model. *in proceedings of 7Th Conference on Intelligent Tutoring Systems,* 668-77. Berlin: Springer-Verlag.

[8] Koedinger, K.R., & Anderson, J.R. (1997). Intelligent Tutoring Goes to School in the Big City. *International Journal of Artificial Intelligence in Education* 8, 30-43

[9] Pintrich, P.R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Reserach* (31), 459-70

[10] Reif, F., & Scott, L.A. (1999). Teaching scientific thinking skills: Students and computers coaching each other. *American Journal of Physics* 67(9), 819-31

[11] Roll, I., Aleven, V., McLaren, B.M., Ryu, E., Baker, R.S., & Koedinger, K.R. (2006) The Help Tutor: Does Metacognitive Feedback Improves Students' Help-Seeking Actions, Skills and Learning? *in proceedings of 8Th International Conference in Intelligent Tutoring Systems,* 360-9. Berlin: Springer Verlag.

[12] Roll, I., Baker, R.S., Aleven, V., McLaren, B.M., & Koedinger, K.R. (2005) Modeling Students' Metacognitive Errors in Two Intelligent Tutoring Systems. *in proceedings of User Modeling 2005,* 379-88. Berlin: Springer-Verlag.

[13] Roll, I., Ryu, E., Sewall, J., Leber, B., McLaren, B.M., Aleven, V., & Koedinger, K.R. (2006) Towards Teaching Metacognition: Supporting Spontaneous Self-Assessment. *in proceedings of 8Th International Conference on Intelligent Tutoring Systems,* 738-40. Berlin: Springer Verlag.

[14] Schworm, S., & Renkl, A. (2002) Learning by solved example problems: Instructional explanations reduce self-explanation activity. *in proceedings of The 24Th Annual Conference of the Cognitive Science Society,* 816-21. Mahwah, NJ: Erlbaum.

[15] Tobias, S., & Everson, H.T. (2002). Knowing what you know and what you don't: further research on metacognitive knowledge monitoring. *College Board Research Report 2002-2003*

[16] White, B.Y., & Frederiksen, J.R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction* 16(1), 3-118

[17] Wood, H.A., & Wood, D.J. (1999). Help seeking, learning, and contingent tutoring. *Computers and Education* 33(2), 153-69

[18] Yudelson, M.V., Medvedeva, O., Legowski, E., Castine, M., Jukic, D., & Crowley, R.S. (2006) Mining Student Learning Data to Develop High Level Pedagogic Strategy in a Medical ITS. *in proceedings of Workshop on Educational Data Mining at AAAI 2006,* Menlo Park, CA: AAAI.