To appear in: Working Notes of the AIED 2001 Workhop "Help Provision And Help Seeking In Interactive Learning Environments."

Investigations into Help Seeking and Learning with a Cognitive Tutor

Vincent Aleven and Kenneth R. Koedinger, *Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA. E-mail: aleven@cs.cmu.edu, koedinger@cmu.edu.*

We'd like to know a little bit about you for our files.
We'd like to help you learn to help yourself.
From: "Mrs. Robinson" by Paul Simon

Abstract In order to develop tutoring systems that strike a better balance between system control and student control, or systems that provide guidance and feedback in order to help students learn help-seeking skills, we need a better understanding of (actual and ideal) help seeking. In this paper, we take modest steps toward this goal: First, we continue to assess students help-seeking behaviour in the PACT Geometry Tutor. Students focus too much on the tutor's bottom out hints, meaning that they are given (all but) the right answer but miss the explanation behind that answer, which is undesirable. Second, we study the relations between help seeking, prior knowledge, and learning outcomes, doing some of the same analyses that were reported in a recent paper by Wood and Wood (1999). Like Wood and Wood, we found that process measures of student-tutor interactions account for considerable amount of variance in post-test scores. Further, there are systematic individual differences in students' help seeking behaviour But unlike Wood and Wood, we found little evidence that help seeking helps learning and we did not find that the influence of help-seeking on learning varies depending on students' prior knowledge.

INTRODUCTION

Many intelligent tutoring systems provide feedback when students enter solutions to problems or intermediate steps but provide context-sensitive hints only when students request them. There are good reasons for designing systems to provide hints only on request, most of which represent a justifiable modesty on the part of system designers. Yet there is also a danger in placing control of hints solely in the hands of students. Both the abuse of help (asking for a hint when in fact the student knows enough to proceed successfully without help) or under-use (not asking for a hint when appropriate) are likely to impede learning or slow it down. For student control to work, students need to make good judgments about their need for help. In reality however students may not be very good at judging their own need for help (Aleven & Koedinger, 2000). Further, there is evidence that students with lower prior knowledge seek help in a less discriminating manner (Nelson-Le Gall, et al., 1990). This means that placing control in students' hands is likely to have the unfortunate effect that those who need help the most are least likely to get it in time.

Thus, there appears to be a need to develop tutoring systems that strike a better dynamic balance between student-controlled help and system-controlled help, volunteering and perhaps even withholding help when appropriate. It would be good also if the system could tailor the amount of student control to the needs of groups of students (Luckin & duBoulay, 1999; Arroyo, et al., 2000) or even the needs of individual students. Ultimately, it is desirable to develop tutoring systems that help students learn not just domain-specific knowledge, as current systems do, but also help students to become better help seekers. But how to accomplish these objectives?

There are a number of reasons why it is difficult to design a system that volunteers help at appropriate times or a system that evaluates whether students' help requests are appropriate. It is not straightforward to determine when a particular step in a particular problem at a particular moment in time is too far beyond a particular student's ability. This requires a more detailed assessment than is usually done in student modeling, yet involves all of the usual difficulties of student modeling, such as the inevitable uncertainty inherent in inferences about a student's knowledge state.

Second and equally importantly, we do not have a complete enough understanding of what constitutes good help seeking behaviour. For example, in general it is good to be persistent in trying to figure out an answer, without asking for a hint. At what point does this persistence become help under-use? Similarly, in general it is good to consult more knowledgeable teachers or peers in order to learn from them. At what point does this tendency become help abuse? We need to identify strategies for good help seeking in sufficient detail so that they can be used in systems to evaluate students' help-seeking behaviour or even, so that they can be explained and taught to students.

Finally, how does students' help-seeking behaviour with a system influence their learning outcome? A recent paper by Wood and Wood reported on the associations between help-seeking, prior domain knowledge, and learning outcomes observed in students' interactions with a tutoring system called QUADRATIC (Wood & Wood, 1999). This system is based on the theory of contingent tutoring, which on the surface, and perhaps on a deeper level as well, has considerable common ground with the design principles behind many other ITSs, including Cognitive Tutors (Anderson et al., 1995; Koedinger et al., 1997). Prior knowledge is an important variable in light of findings reported in the literature that students with higher prior knowledge are better able to assess their need for help (Nelson-Le Gall, et al., 1990). Some of the main findings reported in the Wood and Wood paper were:

- Process measures about students' actions on the tutor (including error frequency and measures of help seeking behaviour) are predictive of students' learning outcomes, above and beyond pre-test scores and measures of prior knowledge.
- The effect of help-seeking behaviour on learning outcome is different depending on students' prior knowledge.
- Finally, there are systematic individual differences in students' approaches to help seeking.

These findings are clearly relevant to designers of ITSs, even if there is still the step of translating such findings into design recommendations.

In the current paper, we sketch a common sense model of desired help-seeking behaviour with the PACT Geometry Tutor. Using data from a classroom study involving the PACT Geometry Tutor, we assess students' help seeking behaviour by loosely comparing it against this model. Further, we study relations between help seeking and learning, performing part of the analyses that Wood and Wood had done. Such an investigation is useful for a number of reasons: First, there is a difference in scale. While QUADRATIC's curriculum comprises about 1-1.5 hours of instruction, the PACT Geometry Tutor addresses a full-year high-school geometry course. The data reported here relate to one of six major units that make up the curriculum, which takes about 8.5 hours of work on the tutor. It is quite possible that certain tendencies in students' help-seeking behaviour do not develop until students work on the tutor over an extended period of time and become more familiar with the tutor. The second reason is that while QUADRATIC and PACT Geometry Tutor are rather similar, they are also different in subtle ways and deal with different domains. This may influence students' help seeking behaviour.

THE PACT GEOMETRY TUTOR

The PACT Geometry Tutor (Aleven, et al, 1999) is an integrated part of a full-year high-school geometry course. Following guidelines of the National Council of Teachers of Mathematics, it emphasizes the use of geometry to solve "real-world problems". In these problems, students are presented with a description of a problem situation plus a diagram abstracted from this situation (some problems involve only a diagram) and are asked to calculate unknown measures. Students are asked also to provide reasons for their (numeric) answers. They can do so by selecting the name of an appropriate geometry theorem or definition from the tutor's on-screen Glossary of geometry knowledge or they can simply type the name (e.g., "Triangle Sum"). The students can use the Glossary freely as they work with the tutor. The Glossary is a source of decontextualized help, similar to many sources that students are likely to encounter in the real world, such as the World-Wide Web. In this paper however we focus not on the Glossary but on the tutor's contextualized hint messages.

When the student enters a solution or intermediate step, the tutor provides feedback indicating whether the step is correct or not. On demand, the tutor provides context-sensitive hints. For each step, between 5 to 8 levels of hints are available, depending on the skill involved. Each hint level provides increasingly more specific advice. When the student first requests a hint, the tutor starts at the first (most general) level. It displays the next level when the student repeats the help request. The hints are designed to communicate a general strategy, namely, when you do not know something, use a resource like the Glossary to look up relevant information and then use that information to solve your problem. The early hints in each sequence typically identify a cue (e.g., that the problem involves a triangle) that can help focus the search for a relevant geometry rule in the Glossary and suggest that the students undertake that search. An intermediate hint highlights the relevant rule in the Glossary. The later hints in each sequence discuss how the rule can be applied to find the targeted angle measure. The hint sequence "bottoms out" by stating an expression describing how to find the unknown quantity, for example, "$m\angle ABC = 180 - m\angle ACB - m\angle CAB$".

Like all Cognitive Tutors (Anderson, et al., 1995; Koedinger, et al., 1997), behind the scenes the PACT Geometry Tutor uses a cognitive model to monitor students as they solve problems. The cognitive model represents the skills targeted in the instruction, represented as a set of production rules. The tutor uses the model to analyze students' activities in terms of the underlying skills (i.e., production rules in the model), through a process called model-tracing. The tutor also maintains a student model, estimating the probability that the student knows each crucial skill in the model. The model is updated, through a process called knowledge-tracing, once for each step—a step is a subgoal in a tutor problem or put differently, a step corresponds to an entry in the tutor's answer sheet. The update of the student model is positive or negative, depending on whether the step was completed successfully by the student, meaning that no errors were made or hints requested. In normal operation, the tutor uses the student model to implement a mastery level criterion, as follows: After the student finishes the set of "required problems" for any given curriculum section, the tutor assigns problems that involve unmastered skills until all skills for the given section are above the mastery level threshold, which has been set to .95. At that point, the student is promoted to the next curriculum section.

Thus, the PACT Geometry Tutor "rewards" the successful completion of each problem-solving step by increasing its estimate of the probability that the student knows the relevant skill. The increase is visible on the computer screen, in the skillmeter window which shows a bar chart with the tutor's estimates of skill mastery. On the other hand, if the student made an error or requested a hint, the visible result is that one of the "skill bars" moves in the wrong direction. The reward for successful completion of problem-solving steps is real in the sense that usually students see it as their goal to have the skills in the skillmeter window be ticked off, which happens once a skill is above the mastery threshold. Further, students are generally aware that the more steps they complete successfully, the fewer problems they will be assigned[1]. In this sense, the PACT Geometry Tutor is different from QUADRATIC, in which there seems to be no such consequence of errors or hints, in other words, errors and hints are free. The reward structure in the PACT Geometry Tutor is motivated to a large degree by the theory behind the tutor's knowledge-tracing algorithm, which treats a help request as a sign that the student does not master the skill that is involved (Corbett & Anderson, 1995). The reward structure may influence students' help-seeking strategies. For example, it may encourage students to be somewhat cautious before entering an answer (probably a good thing). Also, it may keep students from trying to complete problems simply by asking for a hint on each step (definitely a good thing). On the other hand, it may also have the result that students are somewhat reluctant to ask for a hint before trying an answer (which sometimes might not be a good thing). There is some evidence that this kind of "response-contingent reward system" leads to more discriminating help use, at least with younger children, namely, 3rd- and 5th-graders, (Nelson-Le Gall et al., 1990).

If we ignore the tutor's Glossary for the moment, then the desired help-seeking behaviour might be described as follows (this general strategy seems appropriate both for QUADRATIC and the PACT Geometry Tutor):

1. Spend time to think about the step
   a. If it is not at all familiar (e.g., first time for this skill), then ask for help and goto 5.
2. Try step (i.e., enter an answer)
3. If correct, goto 8
4. If error (not a slip) and if it is not clear how to fix the error, ask for help and goto 5.
   a. If you have an idea how to fix the error or slip, try again and goto 3.
5. Spend time to read hint.
   a. If it is clearly not helpful (e.g., says to pursue the goal you just made an error on), then get help and goto 5.
6. Think and try again.
7. Goto 3
8. Done—move on to the next step.

DESIGN OF THE STUDY

The goals of the current study were (1) to study students' strategies for using the tutor's help facilities and (2) to analyse the relations between prior knowledge, measures of the student-tutor interaction (including help seeking), and learning outcomes. We used the data that were collected for a study designed to evaluate the added value of having students explain their problem-solving steps, in the manner described above (Aleven, et al., 1999). In that study, the participants were divided into two conditions, each working with slightly different tutor versions. In the current

[1] In the current study, the tutor controlled the amount of time that students spent on the tutor, rather than promoting students when they reached mastery for all skills. Thus the assumption did not hold that the more steps were completed successfully, without errors or hints, the fewer problems would be assigned. However, students were not aware of this, at least not until they reached mastery for all skills. It seems reasonable to assume that for the most part, they assumed it was business as usual with the tutor.

study, we combine the results from both conditions. We concentrate on data about student-tutor interactions, which are logged automatically by the tutor. We analyse only the data related to the numeric answer steps in the tutor, as half the students did not provide reasons.

Tests

The pre-test and post-test included questions similar to those encountered on the tutor. In these questions, students were presented with a diagram and were asked to find unknown quantities and state reasons for those quantities, by naming a geometry rule that justified the numeric answer given. The tests also included transfer items, in which students were asked to judge whether enough information was given to calculate certain quantities. The students did not encounter questions of this type when they worked on the tutor.

As measure of prior knowledge we used the average of students' scores on two tests taken during the same geometry course prior to the current study. These tests were related to the topics covered in the first and second tutor unit, about area and the Pythagorean theorem, respectively. Knowledge of these topics does not directly help in solving the problems in the tutor's Angles unit, which do not require the use of area formulas or the Pythagorean theorem. (The Angles unit is the curriculum unit to which the data reported in this paper pertain.) Thus, the scores on these tests provide a measure of students' knowledge on very closely related topics. The scores on these tests might serve as a measure of their interest in and aptitude for geometry.

Subjects

The study involved 53 students taking the PACT Geometry course, two periods taught by the same teacher and his assistant. The students were mostly 10^{th}-graders, that is, 15 and 16-year olds. A total of 41 students completed the experiment.

Procedure

The study took place in the course of regular instruction with the PACT Geometry Tutor in a suburban school in the Pittsburgh area. During this course (and hence during the experiment) 50% of classroom time was devoted to solving problems on the tutor, the other 50% to classroom and small-group activities. The students engaged in the following activities:

- two tests measuring prior knowledge, one each after completing the tutor's Area unit and Pythagorean Theorem unit
- a pre-test, prior to their work on the tutor's Angles unit
- work on the tutor's Angles unit - this is one of the six units that make up the full-year curriculum of the tutor. The tutor controlled for time and let students go on to the next unit after 8.5 hours. The students started to work on the tutor's Angles unit when they completed the two prior tutor units, which was at different times for different students.
- a post-test after completing the Angles unit
- classroom instruction, some of which may have taken place before the pre-test.

STUDENTS' HELP SEEKING STRATEGIES

We assessed students' help-seeking behaviour by comparing it (loosely) against the model outlined above. We found that students follow the model of help-seeking behaviour only to a

Table 1: Process measures related to the student-tutor interaction. All measures pertain to numeric answer steps only. Explanation steps (i.e., steps where students gave reasons) are not included.

Process measure	Description	Mean
Success frequency	Percentage of steps that the student got right without making any errors or requesting any hints.	55 ± 18
Help frequency	Percentage of steps on which student asked for help	27 ± 19
Error frequency	Percentage of steps on which student made one or more errors	36 ± 12
Tendency to seek help to avoid error	Number of steps where student used help and did not made an error divided by the number of incorrect steps (i.e., steps where the student used help or made an error)	18 ± 14
Tendency to seek help after an error	Percentage of incorrect steps on which the students used help after an error was made (i.e., when students make an error, how often do they ask for help)	40 ± 20

limited degree, as we had found before (Aleven & Koedinger, 2000). Students got 54% of the numeric answer steps right, without error or help from the tutor. They made errors on 36% of the steps. They used the tutor's help facilities on 29% of the numeric answer steps. Although the model of ideal help-seeking behaviour outlined above does not make quantitative predictions, the overall help frequency seems about right. In particular, it is not too far removed from the error frequency. The tutor data further provide clear evidence that the tutor's hints helped students' performance. Students were correct on 82% of answer attempts that immediately followed a hint, whereas they were correct 62% of the time on steps where their first action was an answer attempt (as opposed to a hint request).

However, students focused largely on the tutor's bottom-out hints, which as described stopped just short of handing them the answer. On 81% of the steps where the students asked for help, they requested to see all hint levels including the bottom out hint. Further, there is evidence of a strategy in which students go through all hint levels rather quickly, until they reach the bottom out hint. They spent one second or less with as much as 68% of the hints at intermediate levels, meaning that their next action after requesting the hint (often this was the next hint request) occurred in less than a second. Finally, students spent an average of 2.3 seconds per intermediate hint, which seems not enough to read and interpret it and decide whether one knows enough to enter the answer.

The finding that students focused on the bottom-our hint was not consistent with our expectation (implicit in the model of desired help-seeking behaviour) that students would need increasingly less detailed help as they became more proficient. The finding is also not consistent with results from an earlier study with a Cognitive Tutor for Lisp Programming, which showed that the average number of hints requested was 1.5 out of 3 hints (Anderson, personal communication—the study but not the specific result mentioned is reported in Harvey & Anderson, 1996). It is still an open question to what extent the abundant use of bottom-out hints reflects a conscious minimum effort strategy on the part of students, or is the result of the students being performance-oriented as opposed to learning-oriented (see e.g., Arbreton, 1998), or means that the hints do not match students' abilities (i.e., that the hints are outside their zone of proximal development). Perhaps the hint sequences have simply become too long.

In order to help students use help more effectively, the current tutor version has been modified (in part by Carnegie Learning, Inc., the company that now markets the software), as follows: First, the tutor volunteers help when a student makes more than two errors on a step (as a way to prevent help under-use). Second, when the student requests a hint, the tutor pauses for two

Table 2: Coefficients *r* of correlation between process measures and prior knowledge, pre-test score, and post-test score. "ns" means that p > .1.

	Prior Knowledge		Pre-Test		Post-Test	
	r	*p*	*r*	*p*	*r*	*p*
Prior knowledge			.08	ns	.38	.02
Pre-Test	.08	ns			.37	.02
Success frequency	.18	ns	.42	.008	.74	.0001
Help frequency	-.17	ns	-.40	.01	-.67	.0001
Error frequency	-.06	ns	-.30	.06	-.66	.0001
Tendency to seek help to avoid error	-.22	ns	-.33	.04	-.40	.01
Tendency to seek help after error	.01	ns	-.27	.10	-.51	.001

seconds before actually displaying the hint (or next hint level), so as to discourage the strategy of clicking on the hint button in rapid succession until the bottom out hint shows up. While we expect some improvement due to these measures, the current investigation is an attempt to find out what can be done beyond these simple heuristics in order to help students to become better help seekers.

PREDICTING LEARNING OUTCOMES

In order to study the relation between help seeking and learning outcomes, we focused on the process measures listed in Table 1. (Some means are different from those reported above because this is the average of the per student average whereas the numbers shown earlier are averages over all steps.) Wood and Wood studied similar measures and considered also the time per operation and latency measures. Studying these important measures remains for future work.

Most process measures correlated significantly with pre-test score but not with prior knowledge (see Table 2). Students who did better on the pre-test had greater success while they worked on the tutor, making fewer errors and requesting fewer hints. Further, they appeared to have less of a tendency to seek help in order to avoid errors and less of a need to use help in order to recover from errors.

Process measures also correlate significantly with the post-test scores. Students who are most successful during training, making fewer errors and requesting fewer hints, tend to do best on the post-test. (Wood and Wood also found this.) Further, students who sought help less often to avoid error and needed help less often to recover from errors, also learned better.

When we partial out prior knowledge and pre-test score, we see the same pattern of results, except that the tendency to seek help to avoid error no longer correlates significantly with the post-test score (see Table 3). Wood and Wood also found the tendency to seek help does not correlate significantly with learning (whether or not prior knowledge is factored out) at least not in the full sample of students.

With the influence of prior knowledge and pre-test scores removed, the need to use help to recover from errors still correlates negatively with learning. On the one hand, it does not seem surprising that students who needed to use help more often to recover from errors would have lower learning outcomes. This would simply be more evidence that those who are in trouble more often during training tend to learn less. In other words, the negative correlation reflects a selection effect. On the other hand, students who used help more frequently had more opportunity to benefit from the help messages—students might acquire greater understanding by reading the explanations in the help messages or might make fewer errors. The negative correlation implies however that any learning advantages due to the more frequent use of help messages were not sufficient to enable those who needed frequent help to catch up. In other words, the advantages of

Table 3: Partial correlation coefficients for correlations between process measures and post-test, when prior domain knowledge and pre-test are partialled out

	All subjects		Subjects with High prior knowledge		Subjects with Low prior knowledge	
	r	*p*	*r*	*p*	*r*	*p*
Success frequency	.70	.0001	.74	.001	.72	.005
Help frequency	-.61	.0001	-.68	.005	-.58	.05
Error frequency	-.66	.0001	-.54	.05	-.80	.0005
Tendency to seek help to avoid error	-.26	ns	-.46	.06	-.12	ns
Tendency to seek help after error	-.54	.005	-.60	.05	-.47	.07

help messages did not offset the selection effect. Maybe this is not surprising in light of students' preoccupation with bottom out hints and more generally the fact that some amount of mathematical reading ability is required to learn from hint messages. Nonetheless, it is somewhat disconcerting not to find any positive correlation between help use and learning outcome, as this suggests that "the rich get richer, the poor get poorer." This kind of trend we would very much like to avoid in the use of intelligent educational software.

We considered whether process measures could help to account for variability in the post-test scores, over and above the variability accounted for by the pre-test scores or the students' prior knowledge. Prior knowledge accounts for 14% of the variability in the post-test scores, pre-test score for 14%. Together, these measures account for 26% of the variance. Of the individual process measures, the frequency of success accounts for most variance in the post-test scores, namely, 54%. When we add prior knowledge, we can account for 62% of the variability in the post-test scores. The two process measures error frequency and the tendency to seek help to avoid error together account for 58% of the variance. These measures together with prior knowledge account for 66% of the variability in the post-test scores. This model accounted for 70% of the variance in the High prior knowledge group and 66% in the Low prior knowledge group. We note that adding the pre-test to any of the four models made them account for almost no extra variance.

These findings are similar to those reported by Wood and Wood, who also found that process measures and prior knowledge together accounted for more variance than either one alone, namely 62%. They found that prior knowledge accounted for more variance than the process measures, whereas in our data, it was very much the other way around. A possible explanation for this difference between the two studies is the time scale of the instruction. As mentioned, in our study students worked about six times as long on the tutor as did students in the Wood and Wood study. The longer the instruction, the more one expects that effects of prior knowledge are overshadowed by those of knowledge acquired during the instruction.

As a practical matter, our results suggests that there is some justification for basing students' grades for work on the tutor in part on process measures such as the frequency of error or the frequency of help requests. Further investigation is needed, however, before we recommend doing this. In the current study the tutor did not use its normal mastery level criterion. It is important to verify that the same results are obtained when the mastery level criterion is used.

INTERACTION OF PROCESS MEASURES WITH PRIOR KNOWLEDGE

We considered whether students' prior knowledge affects the way that help seeking influences learning outcomes. To do so, we ran a series of 2-way ANOVAs with post-test score as the independent variable and prior knowledge and one process measure as regressors. We found no

Table 4: Correlation coefficients for process measures across (a) early, middle, and late opportunities for applying each skill, and (b) curriculum sections.
* p < .0005. For all other correlations, p < .0001

	%Help	%Error	Tendency to seek help to avoid error	Tendency to seek help after error
Steps 1-6 v. steps 7-12	.92	.88	.90	.86
Steps 7-12 v. steps 13-end	.91	.83	.72	.83
Skills of S1 v. skills of S2	.97	.91	.78	.90
Skills of S2 v. skills of S3	.85	.83	.55*	.69

significant interaction with prior knowledge for any of the process measures that were considered. We also studied whether the relation between help seeking and learning differs depending on students' pre-test score, running a second series of ANOVAs with post-test as dependent measure. We found no significant interactions between pre-test score and process measures.

However, a number of interesting tendencies can be observed in the data (see Table 3): First, error frequency is more highly (negatively) correlated with learning outcomes in the Low prior knowledge group than in the High prior knowledge group. This suggests that the negative effect of errors on learning is more pronounced for those with lower prior knowledge. This finding is consistent with, although less strong than, the findings reported by Wood and Wood, who found that in the group with low prior knowledge, but not in the group with high prior knowledge, errors are negatively associated with learning outcome and also found that this difference between groups was statistically significant.

Further, in our data (see Table 3) help frequency is more highly (negatively) correlated with learning outcome in the High prior knowledge group than in the Low group. Also, a tendency to seek help to avoid errors is negatively correlated with learning in the High group but not in the Low group. Again, this is in the same direction as, but less strong than, the findings reported by Wood and Wood, who found that in the low prior knowledge group but not in the high prior knowledge group, help seeking is positively correlated with learning (and again the difference between groups was statistically significant).

CONSISTENCY OF INDIVIDUAL STUDENTS' HELP-SEEKING BEHAVIOUR

We studied whether there were systematic individual differences in students' help-seeking behaviour, as Wood and Wood had found. We used two approaches. First, we looked at the correlations of process measures across early, middle, and late opportunities to apply each skill targeted in the tutor curriculum. (Here "opportunity" means the same as step.) The idea is to see whether students' tendencies remain systematic even as they gain proficiency with a skill. This analysis requires that we map each step in a tutor problem to the skills that are involved in successfully completing the step. This mapping was done on the basis of the tutor's cognitive model. In fact, the tutor produces this information in the normal course of business, as it carries out its model-tracing process, so this information was available in the logs of student-tutor interactions. Specifically, we compared the data from the first 6 opportunities to apply each skill against the data from the 7-12th opportunity against the data from the rest of the opportunities (13th and onwards). We averaged over all skills. The second approach to studying the consistency of help-seeking behaviour was to study correlations of process measures across three groups of skills. We divided up the set of skills into three groups, following the division of the Angles unit of the PACT Geometry Tutor into three main sections[2].

We found that help frequency and error frequency were highly correlated across early, middle, and late opportunities, as well as across skill groups (see Table 4). Students' tendencies to seek help were also significantly correlated across opportunities and skills. .Thus, students who do well with the tutor early on (successfully completing a high percentage of steps) continue to do well later and students who are successful in learning one set of skills also succeed with other skills. More importantly, students who seek help more often on early opportunities to apply a skill also seek help more often on middle and late opportunities. Similarly, students who seek help more often when dealing with skills introduced during the first curriculum section also seek help more often when dealing with skills introduced during the second and third curriculum sections.

Before we conclude that this means that there are systematic individual differences in students' tendencies to seek help, we need to rule out an alternative explanation, namely, that the correlations between early, middle, and late help seeking tendency are solely a consequence of the correlations between early, middle, and late error rate. Although this alternative explanation might seem plausible (for example because errors might trigger help requests), it can actually be ruled out rather easily, since the two measures of help seeking tendency that were used in this study actually factor out the error rate (i.e., have the error count in the denominator, see Table 1)[3]. Therefore, the data support a conclusion that there are systematic individual differences in students' in the way students seek help.

Our findings are consistent with the findings reported by Wood and Wood, although not entirely the same. Wood and Wood also found statistically significant correlations in help seeking between different topics or sessions on the tutor. They ruled out that these correlations would be due solely to systematic individual differences in error rate in a different manner than we did: In their study, correlations in error rate were found to exist only between earlier sessions or topics but not between later sessions or topics. In spite of this difference, our data support the same conclusion.

DISCUSSION AND CONCLUSION

Intelligent tutoring systems often provide feedback on demand. But how good are students at making optimal use of such help facilities? And how does students' help-seeking behaviour relate to learning outcomes? Finally, how does students' prior knowledge influence the relation between help-seeking and learning? The current study presents data related to these questions. First, data from the PACT Geometry Tutor indicate that there is room for improvement in the help-seeking abilities of the target population, 10^{th}-graders taking a geometry course. On the upside, students seemed to use the system's help messages with appropriate frequency and the messages were clearly useful in helping students complete problems. On the downside, students focused too much on the tutor's bottom out hints, which is undesirable, since in doing so students do not pay much attention to the explanation of why the answer is the way it is.

---

[2] The sections are: angles formed by two intersecting lines, angles in triangles, and angles formed when two parallel lines are intersected by a third line.

[3] Consistent with this argument, there is no statistically significant correlation between the error rate and the tendency to seek help to avoid error (r = -.02). On the other hand, there is a statistically significant correlation between the error rate and the tendency to seek help after errors (r = .66, p < .0001). This latter correlation indicates that students who make errors more frequently also ask for help more frequently following an error (perhaps are less able to self-correct errors). Yet, this does not undercut the argument that the error rate is factored out in each of these measures.

Studying the relations between help-seeking, learning and prior knowledge, we found that process measures of student-tutor interactions, including help seeking, are useful in predicting the learning outcomes, even to a far greater degree than pre-test scores or prior knowledge. Further, we found that there are systematic individual differences in students' help seeking behaviour. These two findings were much along the lines of findings by Wood and Wood.

We did not find that measures of help seeking were positively correlated with learning, as Wood and Wood had found in the Low prior knowledge group. Further, we did not find statistically significant interactions in the way prior knowledge and process measures (including help seeking) relate to students' learning outcomes, again in contrast to Wood and Wood. There was some weak evidence in the data to suggest that similar interactions as those reported by Wood and Wood might exist. One possible explanation for the different results might be the larger time scale of the instruction in our study. It is plausible that the influence of prior knowledge might be more pronounced in a study of smaller scale. In the long run, the effect of prior knowledge may be drowned out by that of new knowledge acquired in the course of the instruction.

How can these results be used to design tutoring systems that strike a better balance between system control and student control or even systems that are able "to help students help themselves"—that is, systems that help students to seek help more effectively? It is difficult to take a specification of desired help-seeking behaviour, such as that offered in the current paper, and derive quantitative predictions. What is needed is a theory that yields quantitative predictions. Without such a theory, it is hard to say what kind of correlations one expects to see between process measures such as those studied in the current paper, and the learning outcomes. For example, what kind of relation does one expect to see between the tendency to seek help and learning outcomes? Where Wood and Wood found a positive correlation (in the Low prior knowledge group), we found a negative correlation (in the High prior knowledge group). But what does this say about how judiciously and deliberately students used help and what does this say about the utility of the help messages in the systems that were studied? Is a negative correlation simply the result of a selection effect or is it a sign of ineffective help seeking and/or help messages?

Without a theory that yields quantitative predications, it may be hard to get a grip on these kinds of questions. Getting to such a theory may not be easy. To the extent that a theory of appropriate help seeking behaviour is premised on an individual student's knowledge (or lack there of) relevant to the particular next step at hand, then we need quite fine-grained metrics of both individual students and individual knowledge elements. (See steps 1a, 4a, and 5a in the model of help-seeking behaviour presented in this paper.) Clearly, different students should seek help at different times for different knowledge. Whereas one student may appropriately need help on a step at a particular time, it may not be so (a) for another student on that step, (b) this student on a different step, or even (c) this student on this step at a different time.

Psychometric theories such as item response theory (Hambleton & Swaminathan, 1985) cover two of these issues, that is, they can provide a metric of how "advanced" a student is likely to be and of how difficult a particular step (item) is going to be. However, they do not deal with the issue of time (c). But it is time variability that we are most concerned about because learning is fundamentally about time, namely, change of performance over time. The bottom line is that it is likely that there will be no quick and easy solutions to developing a theory and appropriate empirical tests of desired help-seeking behaviour. That means that there is hard work ahead if we are to make progress. This will be well worth the effort, as help seeking is important to learning of all kinds and in all domains.

Acknowledgements

Dr. Bethany Rittle-Johnson of the CMU PACT Center provided very valuable advice. Any remaining errors are entirely ours. This research is sponsored by an NSF grant to the Center for Interdisciplinary Research on Constructive Learning (CIRCLE), a joint research center located at the University of Pittsburgh and Carnegie Mellon University. We gratefully acknowledge their contributions.

REFERENCES

Aleven, V. & Koedinger, K. R. (2000). Limitations of Student Control: Do Student Know when they need help? In G. Gauthier, C. Frasson, and K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 292-303). Berlin: Springer Verlag.

Aleven, V.,Koedinger, K. R., & Cross, K. (1999). Tutoring Answer Explanation Fosters Learning with Understanding. In S. P. Lajoie & M. Vivet (Eds.), *Artificial Intelligence in Education, Open Learning Environments: New Computational Technologies to Support Learning, Exploration, and Collaboration, proceedings of AIED-99* (pp. 199-206). Amsterdam: IOS Press.

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

Arbreton, A. (1998). Student Goal Orientation and Help-Seeking Strategy Use In S. A. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 95-116). Mahwah, NJ: Erlbaum.

Arroyo, I., Beck, J. E., Woolf, B. P., Beal, C. R., & Schultz, K. (2000). Macroadapting Animalwatch to gender and cognitive differences with respect to hint interactivity and symbolism. In G. Gauthier, C. Frasson, and K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 574-583). Berlin: Springer Verlag.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction*, 4, 253-278.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications.* Boston: Kluwer.

Harvey, L. & Anderson, J. R. (1996). Transfer of declarative knowledge in complex information processing domains. *Human Computer Interaction,* 69-96.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education,* 8, 30-43.

Luckin, R., & du Boulay, B. (1999). Ecolab: The Development and Evaluation of a Vygotskian Design Framework. *International Journal of Artificial Intelligence and Education,* 10, 198-220.

Nelson-Le Gall, S., Kratzer, L., Jones, E., & DeCooke, P. (1990). Children's Self-Assessment of Performance and Task-Related Help Seeking. *Journal of Experimental Child Psychology,* 49, 245-263.

Wood, H. A., & Wood, D. J. (1999). Help seeking, learning and contingent tutoring. *Computers and Education,* 33, 153-169.