On the Benefits of Seeking (and Avoiding) Help in Online Problem-Solving Environments

Ido Roll
Centre for Teaching, Learning, & Technology
University of British Columbia

Ryan S. J. d. Baker
Department of Human Development
Teachers College, Columbia University

Vincent Aleven and Kenneth R. Koedinger
Human-Computer Interaction Institute
Carnegie Mellon University

Seeking the right level of help at the right time can support learning. However, in the context of online problem-solving environments, it is still not entirely clear which help-seeking strategies are desired. We use fine-grained data from 38 high school students who worked with the Geometry Cognitive Tutor for 2 months to better understand the associations between specific help-seeking patterns and learning. We evaluate how students’ help-seeking behaviors on each step in a tutored problem are associated with their success on subsequent steps that require the same skills. Analyzing learning at the skill level allows us to compare different help-seeking patterns within a single student, controlling for between-student variations. Overall, asking for help on challenging steps is associated with productive learning, and overusing help is associated with poorer learning. However, contrary to many help-seeking theories, avoiding help (and failing repeatedly) is associated with better

Correspondence should be addressed to Ido Roll, Centre for Teaching, Learning, & Technology, University of British Columbia, 214-1961 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail: ido.roll@ubc.ca

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hlns.
learning than seeking help on steps for which students have low prior knowledge. These results suggest that novice learners may benefit from engaging in solution attempts before they can make sense of given assistance. Methodological benefits for using local measures of learning are discussed, and comparisons are drawn to other forms of productive failure in problem solving.

Knowing when and how to seek and apply help is an important part of self-regulated learning (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003; Azevedo, Guthrie, & Seibert, 2004; Karabenick & Newman, 2009; Nelson-Le Gall, 1987; Pintrich, 2004; Roll, Aleven, McLaren, & Koedinger, 2011; Tang, Butler, Cartier, Giammarino, & Gagnon, 2006). Classroom-based studies of help seeking offer important insights into factors that affect students’ help-seeking behaviors and outcomes. For example, several studies have demonstrated that obtaining the right level of help at the right time (termed adaptive help seeking) improves learning (Newman, 1994; Ryan, Patrick, & Shim, 2005). However, students in the classroom often avoid asking for help. Among other reasons, it seems that social factors such as embarrassment may deter students from seeking help (Karabenick & Newman, 2009; Ryan, Pintrich, & Midgley, 2001). Also, students often fail to recognize their need for help (Kruger & Dunning, 1999; Roll, Aleven, & Koedinger, 2011a).

Theories of help seeking offer general guidelines for adaptive help-seeking behaviors. For example, a key step in adaptive help seeking is to identify the need for help (Karabenick & Newman, 2009; Nelson-Le Gall, 1987). However, identifying the need for help is an elusive concept. It is a subjective judgment that depends on knowledge level, the actual and perceived difficulty of the problem to be solved, self-efficacy and sense of competence, context, and goal orientation, to name a few (Arbreton, 1998; Nelson-Le Gall, 1987; Ryan et al., 2001). Currently, there is no comprehensive theory that offers an operational definition of “need for help.” For example, what is the desired balance between help requests and errors? Although students should avoid making errors too often, it is probably desirable for students to make some errors, for example, for a student to self-assess his or her ability to solve a certain class of problems (Roll et al., 2011a) or to risk the greater possible learning benefit if unaided performance is successful (cf. Koedinger & Aleven, 2007).

Canonical theories highlight one intuitive principle—the lower the prior knowledge, the higher the need for assistance (cf. Karabenick & Newman, 2009; Wood & Wood, 1999). Wood and Wood (1999) found that errors were detrimental to learning for low-achieving students but not for high-achieving students. This view is supported by the worked example literature, which suggests that novice learners, but not experts, should be given problems together with their solutions (termed the expertise reversal effect; Sweller, 2006). However, other
theories suggest a more complex relationship between knowledge and help. For example, students with low prior knowledge may not be able to make sense of given instruction (Schwartz, Sears, & Chang, 2007). Specifically, the productive failure literature describes a considerable number of instances when early failed attempts are more productive than receiving guidance (Kapur & Bielaczyc, 2012; Mathan & Koedinger, 2005; Needham & Begg, 1991; Roll, Aleven, & Koedinger, 2009; Schwartz & Martin, 2004; Shih, Koedinger, & Scheines, 2010; Westermann & Rummel, 2012). It may be that in some situations, novice students do not have the tools to encode the information given in help in a way that assists their learning.

**HELP SEEKING IN ONLINE PROBLEM-SOLVING ENVIRONMENTS**

Help seeking in online problem-solving environments (such as Khan Academy practice problems or homework sites) has different characteristics from help seeking in traditional classrooms (Aleven, 2013; Aleven et al., 2003). Mainly, when individual students work with computers, students’ help requests may not carry as much of a social price tag as asking a teacher for help in front of the whole class. Also, problem-solving environments often include a simplified help mechanism that asks students to raise a flag but does not require them to explicitly identify their knowledge gap. Instead, the online environment offers hints that are contextual to the relevant problem. Within the current article we look at a specific type of a problem-solving environment called Intelligent Tutoring Systems (ITS; Koedinger & Corbett, 2006; VanLehn, 2006). ITS include models of the learner and the domain. These models allow ITS to estimate students’ level of knowledge of the relevant skills at each moment.

Investigating student help seeking in the context of ITS provides a useful microcosm for examining the effect of help on learning and for identifying productive patterns of help usage. First, ITS keep detailed traces of students’ behaviors. These traces can be used to identify patterns in students’ moment-by-moment choices and offer an opportunity to study help seeking at a much finer grain size (Aleven, Roll, McLaren, & Koedinger, 2010; Perry & Winne, 2006; Roll, Aleven, McLaren, & Koedinger, 2007b). Second, ITS keep a record of students’ performance at the domain level. This information can be analyzed to infer salient factors in learning and help seeking, such as students’ knowledge level (Corbett & Anderson, 1995; VanLehn, 2006). Last, tutored problem-solving environments offer a well-defined context for learning and thus reduce the number of factors that affect students’ help seeking. For example, built-in help-seeking mechanisms offer specific ways in which help can be sought and streamline the form and content of the given help (Aleven et al., 2003, 2010; Roll et al., 2007b; Wood & Wood, 1999).
The need to better understand students’ online help seeking, and the opportunity to study it in well-defined environments, led to an increased interest in studying online help seeking. Arroyo and Woolf (2005) created a statistical model that uses information about students’ help seeking to infer their attitudes toward help. Guo, Beck, and Heffernan (2008) attempted to improve students’ help seeking using explicit instruction, though their scaffolding ended up producing more maladaptive behaviors. Last, metacognitive feedback used by Roll and colleagues (2011) led to transferable improvement in students’ help-seeking behaviors in ITS.

Fewer studies have attempted to associate specific help-seeking patterns and learning in problem-solving environments, and the picture that they draw is inconclusive. Multiple correlational studies have found that using help resources is correlated with learning (Aleven et al., 2003; Beck, Chang, Mostow, & Corbett, 2008; Roll et al., 2011; Wood & Wood, 1999). However, whereas Luckin and Hammerton (2002) found that effective learners seek more and deeper help, Mathew and Mitrović (2008) found that more help corresponds with shallower learning. Shih, Koedinger, and Scheines (2008) estimated student reflection time after receiving detailed hints and before moving to the next problem step and found that longer reflection is associated with greater learning. Most correlational evidence supports a contingent help hypothesis according to which the level of help sought should be negatively correlated with students’ knowledge of the relevant domain (Roll et al., 2011; Wood & Wood, 1999).

To date, very few studies have attempted to establish causal relationships between specific help-seeking patterns and learning in problem-solving environments. Those that have have often found that the correlational results presented here are overly simplified. Renkl and colleagues found that giving hints in the form of instructional explanations in an example-based learning environment improves learning gains (Renkl, 2002; Schworm & Renkl, 2006). However, they also found that hints had a negative effect on students who were prompted to self-explain the examples (Schworm & Renkl, 2006). In previous work we found that students who received feedback on their help-seeking actions improved their subsequent help-seeking behaviors (Roll et al., 2011) yet did not learn better at the domain level (Roll, Aleven, McLaren, & Koedinger, 2007a). In other words, a set of help-seeking patterns was found to correlate with learning, but supporting students in applying it more frequently did not contribute to their domain-level learning gains. Put together, these results suggest that researchers’ understanding of help seeking in problem-solving environments leaves more to be desired.

Our main research question in this article is the following: What are the relationships between specific help-seeking patterns in problem-solving environments and learning? Common to the studies presented previously is their unit of analysis—the student. One limitation to this approach is that additional factors at the student level may be driving both help-seeking patterns and learning outcomes,
contributing to the contradictory findings seen across studies. In the research presented in this article, fine-grained data from ITS are used to reduce the unit of analysis from the student to the specific problem steps solved by each student. This level of analysis allows us to compare different patterns within each student, thus controlling for between-student factors.

Our secondary question seeks to understand the role of help in situations in which students have low prior knowledge. These situations are interesting, as learners with low prior knowledge are the least likely to seek help adaptively (Karabenick & Newman, 2009; Ryan et al., 2001). Does help correlate with learning on problems that require novel skills, or are students better off gaining more experiences, even through failure, before they can make sense of the given scaffolding?

To answer these questions, we first describe the Geometry Cognitive Tutor, a commonly used problem-solving environment that is the context for our investigation. We then describe our metrics for assessing learning and for classifying students’ help-seeking actions to different patterns of behaviors. Last, we detail the results of analyzing help-seeking data from the Geometry Cognitive Tutor and discuss their implications for theories of help seeking in problem-solving environments.

THE GEOMETRY COGNITIVE TUTOR

We focus our investigation of students’ help seeking on data from the Geometry Cognitive Tutor, a commercial ITS for geometry at the high school level. Tutors in the Cognitive Tutor family are used by 350,000 students in U.S. schools each year (Koedinger & Corbett, 2006). Thus, understanding students’ help seeking in this context may have practical implications for learning as well as theoretical contributions to understanding online help seeking. The Geometry Cognitive Tutor is a tutored problem-solving environment. Most of the interaction takes place in the Scenario window (see Figure 1, on the left). The tutor breaks down each problem into subgoals, or steps, and students advance one step at a time. The Geometry Cognitive Tutor gives students immediate feedback on their solution attempts, and students who enter a wrong answer to a step are required to correct it before moving to the next problem.

The Geometry Cognitive Tutor uses the Bayesian knowledge-tracing algorithm to evaluate students’ level of mastery of each of the relevant skills (Corbett & Anderson, 1995). Each problem step in the tutor is mapped onto the skills that are required to solve it. Data from a student’s performance on each skill across problems allow the tutor to infer the student’s level of knowledge of each specific skill. Examples of skills are working with interior angles between parallel lines or
adding up two angles. Students can see their estimated proficiency on each skill in the Skillometer window (top right corner of Figure 1).

The Geometry Cognitive Tutor has two help mechanisms. The first is a searchable knowledge base in the form of a glossary (bottom right corner of Figure 1). Students can browse the glossary and see definitions and examples for the relevant terms and theorems. The second help mechanism is contextual hints. A student can ask for hints by clicking on the “?” button in the Scenario window. The given hints are specific to the problem step that the student is working on at that time. Several levels of hints are available for each problem step, as demonstrated in Table 1. The first few hint levels are principle (or rule) based. Higher levels of hints become increasingly elaborated and instantiate the general rule within the problem at hand (Aleven & Koedinger, 2000). The last hint level essentially gives the answer in order to prevent students from getting stuck and to convert a too-challenging problem to a worked-out example. Each hint request begins with the most general hint, and students are free to browse the different hint levels.
TABLE 1

An Example of the Different Levels of Hints Available to Students in the Version of the Geometry Cognitive Tutor Used in This Study

<table>
<thead>
<tr>
<th>Level</th>
<th>Hint</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>In this problem, you have Triangle OUT. What do you know about triangles that enables you to find the measure of Angle OUT?</td>
</tr>
<tr>
<td>2</td>
<td>Some rules dealing with triangles are highlighted in the Glossary. Which of these reasons is appropriate? You can click on each reason in the Glossary to find out more.</td>
</tr>
<tr>
<td>3</td>
<td>The sum of the measures of the three interior angles of a triangle is 180 degrees. Angle OUT is an interior angle in a triangle. You know the measures of the other two interior angles: Angles UOT and OUT.</td>
</tr>
<tr>
<td>4</td>
<td>Can you write an equation that helps you find the measure of Angle OUT?</td>
</tr>
<tr>
<td>5</td>
<td>The sum of the measures of Angles OUT, UOT, and OTU equals 180°. So you have: ( m\angle OUT + m\angle UOT + m\angle OTU = 180 )</td>
</tr>
<tr>
<td>6</td>
<td>In the previous hint, you saw that: ( m\angle OUT + m\angle UOT + m\angle OTU = 180 ) You can replace ( m\angle UOT ) by 79 and ( m\angle OTU ) by 79. Also, you can use a variable (say, ( x )) instead of ( m\angle OUT ). This gives: ( x + 79 + 79 = 180 )</td>
</tr>
<tr>
<td>7</td>
<td>Find the measure of Angle OUT by solving for ( x ): ( x + 79 + 79 = 180 ) You can use the Equation Solver.</td>
</tr>
</tbody>
</table>

LOCAL MEASURES OF LEARNING

Studies of online help seeking face a chicken-and-egg problem: Do the identified help-seeking patterns increase learning, or do good learners apply the identified help-seeking patterns? One option is that the identified help-seeking patterns cause learning. However, an alternative is that good learners apply productive strategies that are unrelated to help seeking and also apply the identified help-seeking patterns. In the latter case, help-seeking patterns may be associated with learning but not contribute to learning. One of the challenges in answering this question is that of grain size. As long as our measure of learning is at the student level, avoiding the chicken-and-egg problem is impossible, as a common factor at the student level (such as motivation or reading comprehension ability) may determine both learning and the use of certain help-seeking strategies. Instead, comparing alternative strategies (or patterns) within students eliminates between-student variations. Rather than evaluating students’ overall behavior, we can compare different patterns that each student applies to different problem steps. At the risk of abusing the analogy, comparing omelets from different eggs laid by
Local measures of learning. The quality of the solution process on each step is evaluated using students’ performance on the subsequent step that requires the same skill. The same hen can tell us whether the quality of the omelet (or learning) depends on the egg (help-seeking behaviors) or the hen (stable student factors). Thus, the challenge is to differentiate learning from the different patterns of behavior of a single student.

A student’s actions on a certain step affect his or her success on the next step that requires the same skill. Thus, the quality of a student’s help seeking could be evaluated by measuring the improvement from the current problem step to the next one that requires the same skill. For example, the student in Figure 2 first makes an error on the left square, asks for a hint, gets it right, and moves on to the triangle problem. The quality of this process can be evaluated by looking at the student’s performance on the second square problem. A correct answer on the first attempt at the second square problem suggests that the student’s actions on the first square problem were productive. We refer to this approach as a local measure of learning.

What is interesting is that this approach separates successful from productive, as certain behaviors may lead to failed attempts on the current step yet improved learning as measured by performance on future steps, and vice versa. Unlike several studies that focused on evaluating help resources (cf. Baker et al., 2009; Beck et al., 2008), in this article we evaluate help seeking from the student perspective, that is, which help-seeking patterns are most productive.

CLASSIFYING HELP-SEEKING PATTERNS

One of the challenges in evaluating help seeking is the complexity of students’ help-seeking behaviors. Each hint request (or lack thereof) should be interpreted in light of other factors, such as the time that students took before asking for the hint, while reading it, and prior to attempting to solve; the level of hint that they asked to see; their estimated knowledge level; and so on. Thus, our first step was to identify and classify help-seeking patterns in the interaction data. We did so using the Help-Seeking Model (HSM; Aleven, McLaren, Roll, & Koedinger, 2006; Roll et al., 2011). The HSM is a cognitive model applied to a metacognitive domain and it reflects contemporary theories of help seeking. According to the HSM,
students should attempt to solve a step only if they think that they know how to do so. Students who understand the problem but are not sure how to solve it should search the glossary. Students who do not even understand the problem should ask for a hint. Last, students who make an error should reflect on their error. They should then ask for a hint or attempt again, depending on their understanding of their error (see Figure 3). The complete model includes several parameters, such as estimated skill level (e.g., the problem requires a new skill, as estimated by the Geometry Cognitive Tutor), recent history on the same problem step (e.g., the student has already seen one third of the available hints but has not tried to solve the problem step yet), time to take the action (in seconds), and context (e.g., the student has just searched the glossary).

The HSM classifies each student action using 80 production (if/then) rules. This information can be aggregated to classify each student action as one of three types. Desired Help refers to actions that conform to the desired set of behaviors as identified by the model. Help Abuse refers to using help in excess of the estimated need. One form of Help Abuse is often referred to as “clicking through hints.” This behavior describes a student who asks for a hint but, rather than spending enough time to read it, immediately asks for additional hints until he or she reaches the bottom-out hint that conveys the answer. This behavior allows students to obtain the answer to the relevant step without thinking through the material (cf. Baker, Corbett, Koedinger, & Wagner, 2004). Another form of Help Abuse is asking for
help instead of trying, for example, when the student is estimated to know the relevant geometry skill. The third category, Inappropriate Attempts, refers to patterns in which students try hastily. This family includes two main classes: trying to solve when help would probably be more beneficial and guessing rapidly rather than taking time to think about the problem at hand (cf. Baker et al., 2004). Given the fluid nature of these constructs (e.g., sense of familiarity), the model allows for a wide range of behaviors in any situation and only flags about one sixth of the actions as Help Abuse or Inappropriate Attempts (Roll et al., 2011).

For each student on each observed problem step, the Geometry Cognitive Tutor estimates the student’s proficiency level on the relevant geometry skill. Although the estimate of the tutor is a value between 0 and 1, the HSM simplifies this value to one of three levels. High-Skill steps refer to steps on which the Geometry Cognitive Tutor estimates that the student can solve the problem without additional assistance ($P_{\text{know}} > 0.6$). Med-Skill steps refer to steps on which students are estimated to have medium proficiency ($0.4 < P_{\text{know}} < 0.6$). Last, Low-Skill steps refer to steps on which students are estimated to require the most assistance ($P_{\text{know}} < 0.4$). The thresholds between the levels were previously determined by looking to maximize correlation with learning, minimize the rate of actions that are flagged as inappropriate, and minimize sensitivity to small variations in the estimated parameters (Aleven et al., 2006). Note that the HSM prescribes different actions depending on the estimated skill level. For example, students on High-Skill steps are expected to try again after a single error, whereas students on Low-Skill steps are expected to ask for a hint, with all other aspects of the two situations held constant (Aleven, Roll, & Koedinger, 2012). Also, although students on High-Skill steps are not predicted to require more than one third of the available hint levels, the same students on Low-Skill steps are expected to use all available hints.

Construct validity of the HSM was established by correlating its classifications with paper-and-pencil measurements of help seeking (Roll et al., 2011). External validity was established by correlating the models’ classifications across topics and populations (Roll, Baker, Aleven, McLaren, & Koedinger, 2005). Last, predictive validity was established, as the HSM successfully predicts pre-to-post learning gains (Roll et al., 2011).

**METHOD**

**Design**

We evaluate students’ help-seeking behaviors in the Geometry Cognitive Tutor by classifying students’ actions using the HSM and by identifying which patterns of behaviors are associated with learning, contingent on estimated skill level.
Figure 4 Calculating the rates of Desired Help, Help Abuse, and Inappropriate Attempts for each student on each step.

Each problem step in the Geometry Cognitive Tutor is solved through a sequence of attempts, hint requests, and glossary searches. The last action on each step is always the correct answer—the tutor does not let the student proceed until a correct answer is obtained. Because the HSM classifies each of these actions, we can calculate the rate of Desired Help, Help Abuse, and Inappropriate Attempts for each student–step pair (i.e., for each problem step that each student solves). Figure 4 shows an example of this calculation. The first action of the student is to try to solve the step (row 2, a Desired Help action according to the model). The student gets it wrong and tries again, this time too rapidly (row 3, an instance of Inappropriate Attempts). Her second attempt is also incorrect, so she asks for three rapid hints until she receives the bottom-out hint (rows 4–6, instances of Help Abuse). The student reads the bottom-out hint and eventually applies it to solve the problem (row 7, a Desired Help action). Therefore, the rate of Desired Help is 33% (two actions out of six), the rate of Help Abuse is 50% (three actions out of six), and the rate of Inappropriate Attempts is 17% (one action out of six).

A productive solution process improves a student’s ability to independently solve a subsequent problem step that requires the same skill. A correct first attempt on the subsequent step suggests that the student learned the skill; a help request or an error as a first action suggests that the student still lacks understanding of the target skill (as seen in row 12 in Figure 4). Given that students engage in a variety of behaviors in the course of using the Geometry Cognitive Tutor, assessing each solution process locally controls for other factors, such as student-, time-, or problem-specific characteristics.

Note that steps that students got correct on their first attempt were eliminated from the analysis. These steps do not offer an opportunity to study students’ help-seeking strategies, as they last only one action—the correct first attempt.
Participants

We used data from a previous study that used the HSM (Roll et al., 2007a). The analysis in this article includes only data from the 38 students in the control condition of that study, who worked with an unmodified version of the Geometry Cognitive Tutor.

Data were collected over 2 months in a rural vocational high school in western Pennsylvania (3% minorities; 25% mathematical proficiency on standardized state exams; 69% to 31% male-to-female ratio). The students were enrolled in two Geometry Cognitive Tutor classes taught by two different teachers. All students were accustomed to the Geometry Cognitive Tutor and its interface.

Materials and Procedure

Students in the study used an unmodified version of the Geometry Cognitive Tutor with enhanced logging. The enhanced logging kept traces of students’ actions at the transaction level and the HSM rules that applied to each transaction.

Students worked individually with two units in the Geometry Cognitive Tutor: Angles and Quadrilaterals. The units were studied for 1 month each with a month-long break in between (because of standardized testing). Students progressed within each unit at their personal pace.

Analysis

To associate students’ help-seeking behaviors with success on subsequent steps we evaluated the following logistic regression model:

\[
\text{Log Odds}[\text{Student}_i \text{ correct on Opportunity}_{(k+1)} \text{ for Skill}_j] = B_0 + B_{1i} \times \text{Student}_i + B_{2j} \times \text{Skill}_j + B_3 \times \text{HsPatternRate}_{ijk}.
\]

[Student\_i correct on Opportunity\_(k+1) for Skill\_j] receives 1 when the student answers the subsequent step that requires the same skill correctly on the first attempt and 0 if the student makes an error or asks for a hint. B₀ is the intercept (baseline correctness). B₁ᵢ adjusts the baseline per student (and accounts for between-student variations). B₂ⱼ adjusts the baseline per skill (as some skills are harder to learn than others). Last, B₃ evaluates the contribution of the relevant help-seeking pattern. We fit a separate model for each pattern (Desired Help, Help Abuse, and Inappropriate Attempts). We report the slope (B₃), error (SE B), odds ratio (e^B), Z value, and p value from the relevant logistic regressions. B₃ > 0 suggests that there is a positive relationship, such that a higher rate of the specific pattern corresponds with a higher success rate on relevant next steps. A somewhat more intuitive interpretation of the results is suggested by the odds ratio, e^B. An
odds ratio of 1 corresponds to no effect. Odds ratio greater than 1 corresponds to a positive effect, and values between 0 and 1 correspond to a negative effect.

First we describe the overall association between Desired Help and local learning. Then we dive in to evaluate the relationship between all three patterns (Desired Help, Help Abuse, and Inappropriate Attempts) as a function of students’ knowledge level on the relevant skills: high (i.e., approaching expertise), medium, and low (novice).

RESULTS

Overall, students in the study performed 44,008 actions in 25,105 problem steps. After eliminating problem steps that students solved correctly on their first attempt (which therefore offered no meaningful help-seeking behavior), we were left with 25,337 actions in 6,434 problem steps. Of these steps, 4,590 problem steps (71% of the steps that students did not solve correctly on the first attempt) were followed by correct first attempts on the subsequent steps that required the same skill; 1,844 problem steps (29%) were followed by errors or initial hint requests on the subsequent problem steps that required the same skills.

The rate of students’ help-seeking behaviors as a function of their estimated skill level is shown in Table 2.

Overall, the rate of undesired help-seeking behaviors is relatively low on problem steps on which students had medium and high skill levels. At the same time, the rate of undesired help seeking on low-skill-level steps is much higher—38% of students’ actions on steps for which they lacked sufficient knowledge were deemed unproductive (not counting actions in which students were correct on their first attempt, which are omitted from all of the analyses here). The data especially suggest that students try to solve too often on Low-Skill steps. The 28% of actions that were labeled Inappropriate Attempts were all erroneous solution attempts, as

<table>
<thead>
<tr>
<th>Action</th>
<th>All Actions (25,337 Actions)</th>
<th>High-Skill Steps $P_{\text{know}} &gt; 0.6$ (8,863 Actions)</th>
<th>Med-Skill Steps $0.4 &lt; P_{\text{know}} &lt; 0.6$ (7,053 Actions)</th>
<th>Low-Skill Steps $P_{\text{know}} &lt; 0.4$ (9,421 Actions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired Help</td>
<td>83%</td>
<td>91%</td>
<td>90%</td>
<td>62%</td>
</tr>
<tr>
<td>Help Abuse</td>
<td>5%</td>
<td>5%</td>
<td>2%</td>
<td>10%</td>
</tr>
<tr>
<td>Inappropriate Attempts</td>
<td>12%</td>
<td>4%</td>
<td>8%</td>
<td>28%</td>
</tr>
</tbody>
</table>

*Note. Actions on which the student was correct on the first attempt are omitted from these percentages.*
correct attempts were labeled Desired Help. Students possibly engage in a high rate of Inappropriate Attempts on Low-Skill steps because they overestimate their ability (Roll et al., 2011a). Kruger and Dunning (1999) referred to this situation as “unskilled and unaware of it.” It is also important to reiterate that the definitions of Help Abuse and Inappropriate Attempts were different for different skill levels, as described in “Classifying Help-Seeking Patterns.”

**Desired Help-Seeking Behaviors and Learning**

The rate of Desired Help is a significant factor in students’ ability to correctly solve the subsequent problem step requiring the same skill on the first attempt: $B = 0.37$, $SE (B) = 0.15$, $e^B = 1.45$, $Z = 2.52$, $p = .01$ (as a reminder, all successful solution attempts were defined as Desired Help, in addition to desired unsuccessful attempts and desired help requests). The odds ratio ($e^B$) offers a somewhat intuitive interpretation of these values. A student who performs nothing but Desired Help actions has 45% higher odds of getting the next step correct compared with a student whose actions were not classified as desired. Converted to probabilities, this means that a student who has a 50% chance of learning (i.e., getting the next step correct) without any desired help-seeking behavior has a 59% chance of getting the next step correct if all of his or her actions are desired.

However, including all actions in this analysis ignores skill level. Thus, we repeated separate analyses for each of the three groups of skill levels separately, as seen in Table 3.

Engaging in desired help-seeking behaviors is a significant predictor of success on subsequent attempts for students on High- and Med-Skill steps, as shown in Table 3. If a student has a 50% chance of getting the next step right without engaging in desired help-seeking actions, these chances climb to 69% and 73% for High- and Med-Skill steps, respectively. Yet the rate of Desired Help is not

<table>
<thead>
<tr>
<th>Skill Level</th>
<th>$B$</th>
<th>$SE (B)$</th>
<th>$e^B$</th>
<th>$Z$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All steps (83%)</td>
<td>0.37</td>
<td>0.15</td>
<td>1.45</td>
<td>2.52</td>
<td>.01</td>
</tr>
<tr>
<td>High-Skill steps (91%)</td>
<td>0.83</td>
<td>0.34</td>
<td>2.29</td>
<td>2.45</td>
<td>.01</td>
</tr>
<tr>
<td>Med-Skill steps (90%)</td>
<td>1.00</td>
<td>0.29</td>
<td>2.72</td>
<td>3.41</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Low-Skill steps (62%)</td>
<td>−0.37</td>
<td>0.27</td>
<td>0.69</td>
<td>−1.38</td>
<td>.17</td>
</tr>
</tbody>
</table>

*Note.* Reported values are slope ($B$), error in slope ($SE B$), and odds ratio ($e^B$). The rate of Desired Help is in parentheses.

*p < .05. ***$p < .001.
a significant predictor of success on subsequent attempts on problem steps for which students are estimated to have low skill levels.

Types of Help-Seeking Errors

To evaluate the association between undesired behaviors and success rate on subsequent steps we calculated separate regression models for Help Abuse and Inappropriate Attempts. Table 4 shows whether engaging in each of these patterns is associated with success on the subsequent step, as a function of skill level.

As shown in Table 4, Help abuse is clearly associated with poor learning across all skill levels. If a certain student has a 50% chance of learning from a certain step without any instance of Help Abuse, this rate drops to 14%, 7%, and 19% on High-, Med-, and Low-Skill steps, respectively, for a student who only engages in Help Abuse.

The rate of Inappropriate Attempts is a significant predictor of lack of success on the next attempt only on Med-Skill steps but not on High- or Low-Skill steps. In fact, a high rate of Inappropriate Attempts on Low-Skill steps is associated with improved learning, as measured by students’ success on relevant subsequent attempts. Compared with a student who has a 50% chance of learning from a certain step without any inappropriate attempts, a student who engages in Inappropriate Attempts on Med-Skill steps has only a 31% chance of learning, whereas the same behaviors on Low-Skill steps correspond to a 68% chance of learning.

<table>
<thead>
<tr>
<th>Skill Level</th>
<th>B</th>
<th>SE (B)</th>
<th>e^B</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Skill steps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help Abuse (5%)</td>
<td>−1.76***</td>
<td>0.53</td>
<td>0.17</td>
<td>−3.30</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Inappropriate Attempts (4%)</td>
<td>−0.16</td>
<td>0.41</td>
<td>0.85</td>
<td>−0.40</td>
<td>.69</td>
</tr>
<tr>
<td>Med-Skill steps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help Abuse (2%)</td>
<td>−2.48*</td>
<td>0.96</td>
<td>0.08</td>
<td>−2.57</td>
<td>.01</td>
</tr>
<tr>
<td>Inappropriate Attempts (8%)</td>
<td>−0.79**</td>
<td>0.30</td>
<td>0.45</td>
<td>−2.62</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Low-Skill steps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help Abuse (10%)</td>
<td>−1.42***</td>
<td>0.42</td>
<td>0.24</td>
<td>−3.38</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Inappropriate Attempts (27%)</td>
<td>0.79**</td>
<td>0.24</td>
<td>2.20</td>
<td>3.24</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

Note. Average error rates out of all actions in that skill level are shown in parentheses. e^B > 1 implies that a higher error rate corresponds to higher chances of correctly answering the next problem step.

*p < .05. **p < .01. ***p < .001.
DISCUSSION

The data presented here include all problem steps on which students made an error and thus had some room for improvement. In these situations, students could either exploit help (Help Abuse), ask for just enough help (Desired Help), or avoid help (Inappropriate Attempts). Overall, desirable help-seeking actions, as captured by the HSM, correlate with better local learning, as assessed by students’ improvement from the current step to the next step on the same skill. This is predictable, as the HSM was designed to capture adaptive help-seeking behaviors and given that it was previously validated (Aleven et al., 2006; Roll et al., 2011). Also, these results clearly show that a higher rate of Help Abuse is associated with poorer local learning across all skill levels. That is, students who abuse help and avoid engaging with the problem are not likely to succeed on following problem steps involving the same skill. These findings give a partial answer to our main research question: Although some help use is productive, overusing help seems to be detrimental to learning.

However, a closer investigation of students’ help-seeking patterns tells a more complex story. Specifically, students who attempt steps for which their estimated skill level is low may benefit more from failed attempts than from asking for help. Put differently, some assumptions about help appropriateness appear to be least valid in the situations in which students need the most support. Several explanations are suggested further next.

Help-Seeking Patterns on Med- and High-Skill Steps

Desired Help was associated with learning on Med-Skill steps, and Help Abuse and Inappropriate Attempts were associated with lack of learning on these steps. This suggests that adaptive help-seeking behaviors, as defined by the HSM, are most productive on steps that require skills that students have begun learning but have not mastered yet: On the one hand, students have room for improvement; on the other, they know enough to make sense of the given assistance. It seems that the available help in the Geometry Cognitive Tutor matches students’ needs and capabilities on these steps.

Interestingly enough, Inappropriate Attempts were not associated with learning on High-Skill steps. At the same time, Desired Help was associated with increased learning. Desired Help on High-Skill steps included asking for low levels of (rule-based) help after repeated errors. Seeking help in these situations leads to a substantial increase in success on the next opportunity even in High-Skill steps. Abusing help, however, was associated with reduced learning also in High-Skill steps.
Help Seeking on Low-Skill Steps

As was the case on Med- and High-Skill steps, abusing available help was strongly associated with poorer learning also on Low-Skill steps. However, a higher rate of Inappropriate Attempts was significantly associated with a higher success rate on subsequent steps. In fact, avoiding help on Low-Skill steps by attempting (and failing) to solve was found to be more productive for learning than asking for help. This result gives a somewhat surprising answer to our second research question, which focused on ideal help-seeking patterns among novice learners. It is important to understand the roots of the positive relationship between seemingly inappropriate attempts and learning on Low-Skill steps.

The first possible explanation is that the cause and effect are reversed. It is not that students did not learn because they avoided help. Instead they avoided help because they knew these skills better than the cognitive tutor estimated. Perhaps there was a selection effect according to which students avoided help when they were more competent, leading to a correlation between Inappropriate Attempts and learning. However, higher competence suggests that many Inappropriate Attempts should have been correct attempts—but all actions that were labeled Inappropriate Attempts were wrong solution attempts. These are instances of students who try and fail, and it seems that misattribution of skill levels does not explain the observed pattern.

A second potential explanation looks at our outcome variables. While the current article uses a fine-grained assessment of learning, other results suggest that avoiding help is associated with reduced learning in the long term. For example, Baker, Gowda, and Corbett (2011) found evidence indicating that students who avoid help, and keep attempting and failing, do more poorly on later tests of transfer. However, Baker and colleagues looked across all skill levels, and a similar analysis found a similar trend also in the current study. It is only when one looks specifically at Low-Skill steps that the positive relationship between failed attempts and learning is detected. Alternatively, the help mechanisms of the Geometry Cognitive Tutor may assist students in developing conceptual understanding of the material that does not manifest in immediate procedural improvements.

A third, and more likely, explanation suggests that some of the help that is offered by the Geometry Cognitive Tutor may not be helpful in all situations. Note that the same help resources that did not aid students on Low-Skill steps were helpful to the same students on Med-Skill steps. On steps on which students had basic knowledge, using help when needed contributed to learning more than floundering. It is only when students had a low skill level that using help was found to be unhelpful. Other forms of help (such as worked examples) may be more suitable for students in these situations. It is important to remember that students who avoided help still received some scaffolding, as the system flagged
their errors on their failed solution attempts. Although students may have chosen not to receive assistance, the system gave them clear visual cues that their answers were incorrect.

Productive Failure in Geometry Problem Solving

The finding that failed attempts are preferable on Low-Skill steps and help requests are preferable on Med-Skill steps resembles other activities in which failed attempts create a “time for telling” (Schwartz & Bransford, 1998). Early failures may provide valuable learning experiences, even though they do not lead to immediate successful completion of the problem (Kapur, 2008; Roll et al., 2009, 2011b; Schwartz & Martin, 2004; Westermann & Rummel, 2012). It may be that students on Low-Skill steps benefit more from their own struggle rather than from learning from instructional explanations. The productive failure literature further demonstrates that students who attempt multiple solution approaches, albeit failed approaches, learn more than students who attempt fewer solution approaches (Kapur, 2012; Wiedmann, Leach, Rummel, & Wiley, 2012). Thus, learning from repeated errors may be more significant than learning from a single error.

We may be underestimating the dependency of learning from help on prior experiences. In order to learn from given explanations, students should interpret these correctly and extract the general principles in a manner that is transferable to subsequent problem steps. Perhaps students who asked for help on Low-Skill steps could not make sense of it, or perhaps they were able to apply it to the problem at hand but not to generalize across multiple problems. Within the context of problem-solving environments, previous research has found that detailed instructional feedback may improve students’ performance on problems without improving their learning as assessed at posttest (Anderson, Corbett, Koedinger, & Pelletier, 1995; McKendree, 1990). Furthermore, at times, instructional explanations may hamper learning by deterring students from self-explaining their own answers (Mathan & Koedinger, 2005; Schworm & Renkl, 2006; Shih et al., 2010). Thus, attempts on Low-Skill steps may give students valuable experiences with which they can interpret and ground subsequent instruction (Schwartz et al., 2007). At the same time, students who struggle on Med- and High-Skill steps have already had these expository experiences and are ready to be told (Schwartz & Bransford, 1998). Additional research is required to better understand the relationship between actions labeled as Inappropriate Attempts and learning on Low-Skill steps. Were students able to eventually solve these steps by themselves (hinting at a potential generation effect; McNamara & Healy, 2000), or was help more helpful after several failed attempts? Simply put, which failed attempts are productive?
On the Use of Fine-Grained Data

The results described in this article demonstrate how the use of fine-grained data can inform theories of help seeking in problem-solving environments. The HSM uses data at the transaction level to evaluate learning and help-seeking behaviors. Although the use of fine-grained data is common with machine-learned models (cf. Baker & Yacef, 2009), the work described here uses a rational cognitive model (i.e., an interpretable computational model that was built top-down, grounded in theories of help seeking). Such concrete theorizing allows us to evaluate complex patterns of behavior in a hypothesis-driven process and facilitates more precise and clear interpretations of results.

The use of fine-grained data allows us to control for between-student variations. Yet within-student factors may affect the findings. In particular, there may be a selection effect that is not captured by the analysis. For example, we do not know what causes students to avoid, use, or abuse help. Experimental approaches are needed to close the loop.

Limitations and Future Work

The analysis presented in this article identifies one dimension that affects the usefulness of help, namely, students’ level of knowledge of the relevant domain-level skills. Additional factors such as motivation, type of help given, and type of task should also be investigated.

Another limitation of this work is its scope. The analysis was done with one sample of students on two geometry topics within the same tutoring environment. Note that the Geometry Cognitive Tutor is an ITS, and as such it maintains an estimate of students’ knowledge level on each skill. The validity of the findings in other kinds of online environments has yet to be evaluated. There are good reasons to assume that the results generalize to other problem-solving environments with similar help mechanisms. Our analysis evaluated learning from static problem-specific hints (static in the sense that the hints do not adapt to students’ levels). A large variety of online environments offer similar help mechanisms, such as Kahn Academy tutorials (http://www.khanacademy.org), Mastering Physics (http://www.masteringphysics.com), or ASSISTments (http://www.asssitments.org). Given the similarity in the behavior of the help mechanisms across these environments, it is likely that our findings also apply to these environments. It is still unclear, though, to what extent our findings apply to interactive learning environments with other help mechanisms such as learning resources (as in Coursera’s Massive Open Online Courses) or conversational agents (as in River City). Evaluating help seeking with data from a variety of domains, problem-solving environments, and student populations is warranted.
Last, more research is needed to better understand the relationship between local learning and long-term gains on assessments of near and far transfer. Analysis at the skill level should establish that successful local learning correlates with pre-to-post gains on the same skills and should highlight areas in which these two measures differ.

**CONCLUSION**

The work presented in this article uses fine-grained data in order to associate specific help-seeking patterns with learning in a manner that is not possible with more coarse measures (such as learning gains from pre- to posttests). Specifically, we assessed learning locally by measuring students’ ability to correctly solve a new step that requires the same skill as a current step. Our results reaffirm the important role that help seeking plays in learning in highly structured learning environments such as tutored problem-solving environments.

The analysis presented here suggests several conclusions regarding students’ help-seeking behaviors. First, overusing help is associated with lower learning gains across all skill levels. Second, it seems that on steps on which students are proficient (High-Skill steps), avoiding help has little effect on learning. This makes sense, as students should already be knowledgeable enough to solve these steps without additional assistance. On steps on which students are somewhat familiar with the required skill (Med-Skill steps), avoiding help is associated with poor learning, and it is important that students use available help when they struggle. These steps have two important features that make help useful: First, students are in need of help. Second, they have sufficient knowledge to learn from it. Most interesting is that on steps for which students lack basic knowledge (Low-Skill steps), failed attempts are more productive than seeking help. Two complementary explanations can account for this finding. First, learning from given help is probably harder than it seems. The process of understanding, applying, and generalizing from given help may be quite complicated and requires some prior knowledge. Second, we may underestimate the value of failed attempts. Attempting to generate an answer, applying it to the problem step, and reflecting on its failure may support students in acquiring the knowledge that is required to learn from given instruction.

Overall, the research described here achieves two main goals. First, it makes theoretical contributions by associating different patterns of help seeking with learning. Understanding how students learn by seeking (or avoiding) help is important, especially given the scarce evidence for the effect of help. Second, it demonstrates the potential of using analytical methods based on local measures of learning to study complex behaviors. Applying a similar methodology to study other constructs of self-regulated learning can help experts understand and support students in becoming better independent learners.
FUNDING

This work was supported by the Pittsburgh Science of Learning Center, which is supported by the National Science Foundation (SBE-0836012), and by the University of British Columbia through the Carl Wieman Science Education Initiative.

REFERENCES


Baker, R. S. J. d., De Carvalho, A. M. J. B., Raspat, J., Azeven, V., Corbett, A. T., & Koedinger, K. R. (2009). Educational software features that encourage and discourage “gaming the system.” In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. Graesser (Eds.), Proceedings of the 14th international


& B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational
applications* (pp. 283–301). Hillsdale, NJ: Erlbaum.

Perry, N. E., & Winne, P. H. (2006). Learning from learning kits: gStudy traces of students’
self-regulated engagements with computerized content. *Educational Psychology Review, 18*(3),
211–228. doi:10.1007/s10648-006-9014-3


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

Roll, I., Aleven, V., & Koedinger, K. R. (2009). Helping students know “further”—Increasing the
flexibility of students’ knowledge using symbolic invention tasks. In N. A. Taatgen & H. van Rijn
(Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1169–1174).
Austin, TX: Cognitive Science Society.

Roll, I., Aleven, V., & Koedinger, K. R. (2011a). Metacognitive practice makes perfect: Improving stu-
dents’ self-assessment skills with an intelligent tutoring system. In G. Biswas (Ed.), *Proceedings of
the international conference on artificial intelligence in education* (pp. 288–295). Berlin, Germany:
Springer Verlag.

Roll, I., Aleven, V., & Koedinger, K. R. (2011b). Outcomes and mechanisms of transfer in inven-
tion activities. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual
conference of the Cognitive Science Society* (pp. 2824–2829). Austin, TX: Cognitive Science
Society.

Searching for the secret sauce of metacognitive tutoring. In R. Luckin, K. R. Koedinger, & J.
Greer (Eds.), *Proceedings of the international conference on artificial intelligence in education*

Applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning,
2*(2), 125–140.

skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*,
267–280.

metacognitive errors in two intelligent tutoring systems. In L. Ardissono, P. Brna, & A. Mitrovic
(Eds.), *Proceedings of the 10th international conference on user modeling* (pp. 379–388). Berlin, Germany: Springer-Verlag.

teacher as having avoidant, appropriate, or dependent help-seeking tendencies in the classroom.
*Journal of Educational Psychology, 97*(2), 275–285.

Ryan, A. M., Pintrich, P. R., & Midgley, C. (2001). Avoiding seeking help in the classroom: Who and


encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2),
129–184.


