Automated, Unobtrusive, Action-by-Action Assessment of Self-Regulation During Learning With an Intelligent Tutoring System

Vincent Aleven

Human–Computer Interaction Institute
Carnegie Mellon University

Ido Roll

Carl Wieman Science Education Initiative
University of British Columbia, Vancouver, Canada

Bruce M. McLaren and Kenneth R. Koedinger

Human–Computer Interaction Institute
Carnegie Mellon University

Assessment of students' self-regulated learning (SRL) requires a method for evaluating whether observed actions are appropriate acts of self-regulation in the specific learning context in which they occur. We review research that has resulted in an automated method for context-sensitive assessment of a specific SRL strategy, help seeking while working with an intelligent tutoring system. The method relies on a computer-executable model of the targeted SRL strategy. The method was validated by showing that it converges with other measures of help seeking. Automated feedback on help seeking driven by this method led to a lasting improvement in students' help-seeking behavior, although not in domain-specific learning. The method is unobtrusive, is temporally fine-grained, and can be applied on a large scale and over extended periods. The approach could be applied to other SRL strategies besides help seeking.

Recently, there has been a surge of interest in studying students' self-regulatory learning (SRL) processes in computer-based learning environments (CBLEs), such as hypermedia (Azevedo, Cromley, & Seibert, 2004; Brusilovsky, 2001; Dillon & Gabbard, 1998) or inquiry learning software (de Jong & van Joolingen, 1998; Edelson, Gordin, & Pea, 1999; Kuhn, Black, Keselman, & Kaplan, 2000). CBLEs are an attractive platform for investigating self-regulated learning for a number of reasons (Azevedo, 2007; Azevedo et al., 2004; Biswas, Leelawong, Schwartz, Vye, & the Teachable Agents Group at Vanderbilt, 2005; Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Wagster, Tan, Wu, Biswas, & Schwartz, 2007). First, these types of learning environments are now in widespread use in many real-world educational settings. Further, they often place high demands

on students' self-regulatory abilities. Therefore, relations between self-regulation and learning may be especially exposed in these environments. Finally, CBLEs can be instrumented to record data about students' observable actions as they work with the software (Winne & Perry, 2000; Zimmerman, 2008), such as the number of times that a student takes a note with a note-taking tool embedded in the software, moves to a new or already-seen hypermedia page, or requests a hint from an intelligent tutoring system. These environments can thus be used rather conveniently to collect data about student behavior in an unobtrusive manner (i.e., without disrupting natural student work) on a large scale, in significant detail, and over extended periods. With respect to the special issue theme, the data collected with these systems tend to be fine-grained event data. They thereby help support a view of SRL as a sequence of events. By contrast, other popular methods for studying SRL, such as surveys (Zimmerman, 2008), provide data about students' (self-reported) broad tendencies to use or not use particular SRL strategies or processes. These

Correspondence should be addressed to Vincent Aleven, Human–Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15217. E-mail: aleven@cs.cmu.edu

data almost inevitably lack the grain size to support an event-based view of SRL. It is hard if not impossible to reconstruct, from survey data, with any level of detail, actual sequences of mental and observable events that took place during an extended learning episode.

To take advantage of the event-based data produced by CBLEs, it is necessary to interpret the data in terms of SRL processes. One way to do so is to count specific types of observable actions that are supported by the CBLE, especially actions that may directly reflect certain SRL strategies, such as use of a note-taking tool, a highlighting tool, a summarization tool, and so on. This line of work (dubbed "trace analysis") has shown, for example, that students' retrospective self-report about the frequency of their use of SRL strategies tends not to be highly accurate when compared to actual strategy use (e.g., Winne & Jamieson-Noell, 2002), lending support to event-based approaches to studying SRL.

The current article reviews a line of research that is akin to trace analysis in that it focuses on interpreting data about observable behaviors in a CBLE but is also substantially different in that it produced a method that assesses, automatically and action-by-action, whether the student uses a strategy in an appropriate manner. The reviewed research, which was carried out over a period of approximately 8 years, focused on developing an automated assessment method for one particular SRL strategy, help seeking, which has been identified as a key strategy in theories of SRL (e.g., Pintrich, 2004) and has been studied extensively in social learning situations (e.g., Karabenick & Newman, 2006; Newman, 2008; Zusho, Karabenick, Bonney, & Sims, 2007). The research focused on help seeking with an intelligent tutoring system (ITS), a type of software that provides detailed guidance to students as they learn a complex cognitive skill through problem-solving practice (Anderson, Corbett, Koedinger, & Pelletier, 1995; Koedinger & Aleven, 2007; Koedinger & Corbett, 2006; VanLehn, 2006; Woolf, 2009). Students' help-seeking behavior is believed to be a strong influence on learning outcomes with ITSs, yet there is significant evidence that many students do not seek help effectively (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003; Wood & Wood, 1999). We were interested in developing an automated method for assessing student help seeking within this type of CBLE for two main reasons. First, an automated assessment method would enable researchers to study students' help-seeking behavior in significant detail over extended periods. Second, equipped with such a method, the system would be able to provide feedback on students' help-seeking behavior and thereby potentially improve student help seeking and learning.

A significant challenge in devising a method to assess strategy use is that one must take into account not just the actions taken but also the context in which actions are performed. Observable actions such as re-reading, note taking, summarizing, or asking for help are highly appropriate in some contexts but less so in others. Asking for help, for example, when done at the right time, can be a highly adaptive learning behavior that helps learners on their path to independent competence (Aleven et al., 2003; Karabenick, 1998; Karabenick & Newman, 2006; Wood & Wood, 1999). However, when students frequently seek help at inappropriate times, or use help in an executive manner (i.e., as a way to get answers, rather than as an opportunity to learn), help seeking may become an obstacle to learning (Nelson-LeGall, 1985). To assess the quality of students' help-seeking behavior within a given learning environment, therefore, one must go beyond mere counts of the number of help requests, or the number of problems or problem steps finished without help. One must distinguish between situations in which seeking help (in a given manner) is appropriate and likely to be conducive to learning versus situations in which it is not. One must also distinguish between situations in which not using help is appropriate versus situations in which nonuse should be considered help avoidance. Factors to be taken into consideration may include the detailed knowledge components involved in the current step in the current problem, how familiar the student is with these knowledge components, the numbers of errors that the student has made on the current step, the amount and the nature of the help that the student has already requested pertaining to this step, and the time spent on the last request. When assessing use of other SRL strategies, a different (though perhaps overlapping) set of factors may be relevant. Current theories of SRL, however, do not typically provide the requisite level of detail about the conditions under which strategies should be used. A key challenge, therefore, in devising an automated assessment method lies in defining these conditions with great specificity.

To address this challenge, we created a novel assessment method that relies on a computer-executable model of help seeking, which we built by applying traditional cognitive modeling techniques (e.g., Lovett, 1998; Newell & Simon, 1972). The model captures the conditions under which seeking help is a productive behavior and conditions under which it is not. Thus, it is sensitive to the learning context in which the student seeks help, in the manner just discussed. Embedded in the learning software, the model is used as a norm to interpret and assess students' help-seeking behavior. This interpretive process, called "model tracing" (Anderson et al., 1995), in which student behavior in a given situation is compared against a model's behavior in the same situation, is a traditional and widespread technique for ITSs. Prior to the research reported in this article, it had been applied to assess students' domain-specific skills but not their self-regulatory skills. The assessment method is unobtrusive and is applicable on a large scale. It quintessentially reflects a viewpoint of SRL as sequence of events. It may generalize to other SRL strategies, although the testing of this generalization claim is left for future work.

In this article, we review the model-tracing approach to assessing help seeking. We discuss the executable model of help seeking that drives the method. We review evidence for the validity of this method—in particular, its correlation to

other measures of help seeking. We also review the results of a classroom experiment in which the method was used to provide detailed feedback on help seeking within an ITS. We then discuss strengths and limitations of the method and reflect on some of the benefits and challenges of viewing SRL as a sequence of events.

CONTEXT: INTELLIGENT TUTORING SYSTEMS

Intelligent tutoring systems (Koedinger & Aleven, 2007; Schofield, 1995; VanLehn, 2006; Woolf, 2009) provide step-by-step guidance as students learn a complex cognitive skill. They support practice on recurrent problem types (e.g., algebraic equation solving, quantitative physics, computer programming, geometry proof or problem solving). They typically provide (a) a user interface carefully designed to make thinking visible by prompting for intermediate reasoning steps; (b) feedback on the correctness of these steps, not just the final solution to a problem; (c) context-sensitive next-step hints, usually made available at the student's request; and (d) individualized problem selection, based on a detailed assessment of each student's problem-solving skill. ITSs have been shown to improve student learning in real educational settings, compared to other forms of instruction (Anderson et al., 1995; Koedinger & Aleven, 2007; Koedinger, Anderson, Hadley, & Mark, 1997; Mitrovic & Ohlsson, 1999; VanLehn et al., 2005). The research reviewed in this article used the Geometry Cognitive Tutor (see Figure 1), one of a family of ITSs grounded in cognitive theory (Anderson & Lebière, 1998) and cognitive task analysis (Anderson et al., 1995; Koedinger & Corbett, 2006). The Geometry Cognitive Tutor is an integrated part of a full-year high school geometry course that at the time of this writing is in use in approximately 600 schools nationwide.

As is typical of ITSs, at every step along the way to solving a tutor problem, a student can request a hint from the Geometry Cognitive Tutor. Typically, multiple levels of context-specific hints are available, which progress from more general to more specific advice. It is up to the student when to request hints and how many hint levels to view before going back to problem solving. The tutor's hints are tailored to the specific step and the student's problem-solving approach. Earlier hint levels point out an appropriate step to work on next, and point to features of the problem that suggest which geometry theorem to use (e.g., the fact that the problem involves parallel lines or an isosceles triangle). The intermediate hint levels typically describe a theorem and discuss how to apply it. The last hint level (which we call the "bottom-out hint") typically provides the answer. The Geometry Cognitive Tutor also provides a second source of help, namely, an online glossary with theorems and definitions, illustrated with examples (see Figure 1, top right). This source of information is more like real-world sources of help (e.g., math sites on

the Internet) than the tutor's hints. Practice with the glossary may transfer to the use of these real-world sources, a primary reason for including this tool in the tutor.

Each Cognitive Tutor has an executable *cognitive model* of student problem solving, expressed as production rules (Anderson, 1993). The model is an explicit representation of the knowledge components targeted in the instruction. To give a sense for the grain size of these knowledge components, one such component in the geometry tutor's model captures application of the isosceles triangle theorem to compute the measure of one of the base angles in an isosceles triangle, given the measure of the other base angle. The tutor uses the model to assess students' problem-solving actions by comparing them against the actions that the model would take in the same situation. As mentioned, this process is called model tracing (Anderson et al., 1995). Further, using a Bayesian algorithm, the tutor estimates the probability that the student knows each knowledge component captured in its domain model, on the basis of her history of successes and failures on steps involving each knowledge component. It uses these estimates as the basis for individualized sequencing of problems (Corbett & Anderson, 1995; Corbett, McLaughlin, & Scarpinatto, 2000). As discussed next, the model of help seeking also uses these estimates. In the model, the decision to seek help depends among other things on the familiarity of the given step, that is, the student's mastery of the knowledge involved.

A MODEL OF STUDENT HELP SEEKING

It is becoming increasingly evident that self-regulation plays an important role in learning with ITSs (e.g., Koedinger, Aleven, Roll, & Baker, 2009). Examples of recurrent self-regulatory choices that students make when working in these environments are whether to seek help or persist in trying to figure out problem solutions without explanatory help (Aleven et al., 2003), whether to try to game the system (i.e., exploit regularities in the software to get to answers to problems without expending the mental effort required to understand these answers; Baker et al., 2008; Walonoski & Heffernan, 2006), whether to spontaneously self-explain problem-related information even when not prompted by the software (Shih, Koedinger, & Scheines, 2008), and whether to be on-task or engage in off-task behavior (Baker, 2007).

As mentioned, the research reviewed in the current article focused on one particular SRL strategy, help seeking. Prior research had shown that appropriate help-seeking behavior in an ITS or other CBLE can be a associated with better learning outcomes but also that students often use help facilities in ways not conducive to learning (e.g., Aleven et al., 2003; Wood & Wood, 1999). Early on in the research reviewed in the current article, we observed that many students tend to use the tutor's on-demand help facilities in an executive manner, focused on getting answers rather than understanding them
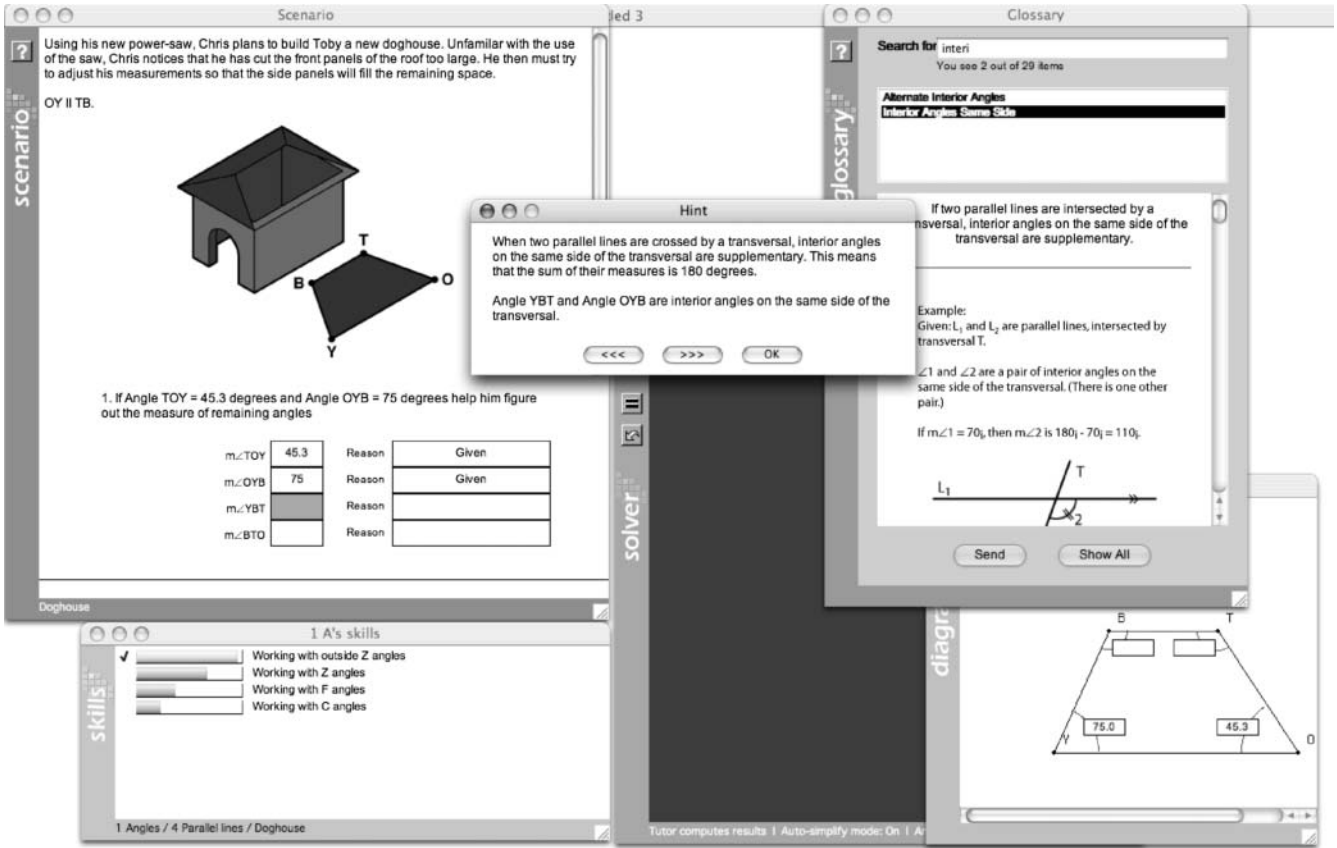
FIGURE 1   The Geometry Cognitive Tutor was used as platform for the research described in this article.

(Aleven & Koedinger, 2000). Similarly, students often did not request help in situations in which they could obviously benefit from it, such as after making multiple errors on a step. We hypothesized that giving students feedback on their help-seeking behavior with the ITS would help them to become better help seekers and better learners. To provide this feedback we created a context-sensitive assessment method.

As a first step in creating such a method, we built a computer model of help seeking with an ITS (Aleven, McLaren, Roll, & Koedinger, 2006). We built the model using a cognitive modeling approach. As we developed the model, we ran it off-line against tutor log data to make sure that productive and unproductive help-seeking behavior were defined in the model in ways that correlated positively and negatively, respectively, with students' learning gains with the tutor. The flow chart in Figure 2 shows the main decision points (depicted as rhombi) and other mental steps (depicted as rectangles) that make up recommended help-seeking behavior. According to the model, students should think about a step in a tutor problem before deciding whether to try the step or request a hint. If the step is familiar and they have a sense of what to do, they should try the step. If the step is not familiar, on the other hand, they should request a hint. Otherwise, students should consult the tutor's glossary. When students have made

an error and it is not clear how to fix the error, they should request a hint. When reading hints, they should spend an appropriate amount of time trying to process and understand the hint before deciding whether to request the next hint level or to try the step. Key decision points in the model capture feeling of knowing (FOK) and judgment of learning (JOL) processes, key SRL processes (Azevedo et al., 2004; Bembenutty, 2009). If students feel they learned or know the knowledge required for a step, then they do not need to (and
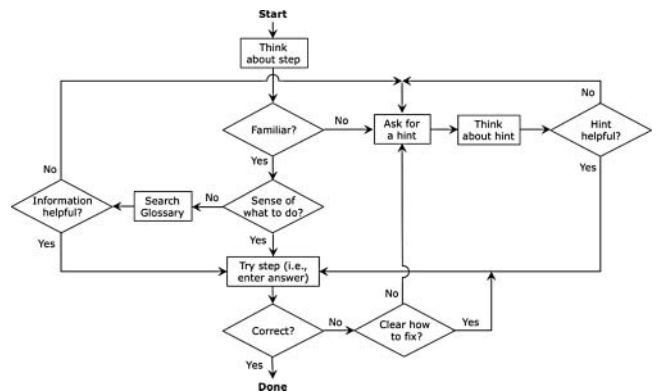

FIGURE 2   The help-seeking model used to assess students' help-seeking behavior.

should not) ask for help or go to the glossary. If they judged that they learned from a hint, then they should not ask for the next hint level, but should try the step. These judgments are simulated in the model based on the tutor's assessment of student geometry knowledge, which, as mentioned, the tutor maintains as part of its regular operation. For example, an unfamiliar step is one for which the estimated probability that the student masters the relevant geometry knowledge component is low (less than 0.4).

The model is context sensitive in a number of ways. First, the recommended behavior depends on the particular knowledge component involved in the given step as well as on the student's familiarity with that knowledge component. When a step is familiar, students are expected to try the step without first requesting a hint, even if one does not necessarily expect flawless execution on familiar steps. On familiar steps, students are expected to ask for hints only when making errors that do not have clear fixes. When a step is unfamiliar, on the other hand, students are expected to ask for a hint immediately, even before trying the step. The model is context sensitive also in that the appropriateness of requesting a hint (or the next hint level) depends on the number of hints seen and errors made on the given step. After multiple errors, students should request help, even if the step is (supposedly) familiar. After multiple hint levels, students should request the next hint level only when the step is unfamiliar. When a step is familiar, they are expected to need only a few of the initial (general) help levels to figure out how to proceed (and the more familiar the step, the fewer levels). As a final way in which the model is context sensitive, the requisite thinking time before an action is taken into account. Hasty actions are considered unproductive, but what is deemed hasty depends on the familiarity of the step.

In addition to modeling recommended help-seeking behavior, the model captures undesirable help-seeking behavior. It contains many detailed types of help-seeking errors belonging to three broad categories: help abuse (e.g., hasty hint requests such as clicking through hint sequences to reach the last level, which typically gives the answer), help avoidance (e.g., trying unfamiliar steps without using hints, or not requesting help after multiple errors on a step), and try-step abuse (e.g., very quick answers on unfamiliar steps). These error types are defined in a manner that is sensitive to context (e.g., whether or not an attempt at trying a step is viewed as help avoidance depends among other things on the familiarity of the step).

The model was implemented as 83 production rules, 35 of which represent productive help-seeking behavior (Winne & Hadwin, 1998) and 48 of which model help-seeking behavior that is undesirable. The model bears a resemblance to models of help seeking found in the social psychology literature (Gross & McMullen, 1983; Nelson-LeGall, 1981; Newman, 1998), as discussed elsewhere (Aleven et al., 2006; Aleven et al., 2003), but is not as comprehensive and focuses less on social aspects of help seeking. Whereas other models of

SRL described in the literature (Pintrich, 2000, 2004; Winne & Hadwin, 1998; Zimmerman, 2000) cover a wide range of SRL processes and strategies, the current model covers a single strategy, although with much greater detail and specificity. In contrast to other models, the current model has a detailed error taxonomy. It is also executable on a computer, for which other models lack the specifics. As a final comment, the model strongly reflects a view of SRL as a sequence of events. The model defines productive help-seeking behavior in terms of observable events and mental events, according to the flow chart of Figure 2.

VALIDATION OF THE MODEL-TRACING METHOD FOR ASSESSING HELP SEEKING

To use the model for automated, real-time assessment and feedback, it was embedded in the ITS (the Geometry Cognitive Tutor), using the same model-tracing algorithm that the software uses to interpret and assess students' domain-level behaviors (Anderson et al., 1995). Thus instrumented, the Geometry Cognitive Tutor assessed each student action (an attempt at solving a step in a geometry problem or a hint request) to determine whether the action constituted desirable help-seeking behavior in the given context or whether it should be viewed as undesirable help-seeking behavior. Specifically, if the action conformed to the model of help-seeking behavior, represented in Figure 2, it was deemed desirable. If the action matched undesirable help-seeking behavior captured in the model, or did not match the model, it was deemed undesirable.

The validity of the model-tracing method for assessing students' help-seeking behavior was tested in a classroom study (Roll, Aleven, McLaren, & Koedinger, in press; Roll et al., 2006), which focused on testing whether the method correlates with other measures of student help seeking. The experiment involved 30 students in two classrooms in a suburban high school in which the Geometry Cognitive Tutor was used in the regular geometry course. As is customary in Cognitive Tutor courses, the students used the tutor for two class periods each week, for a total of six 45-min sessions. During these sessions, the tutor assessed students' help-seeking behavior on an action-by-action basis, using the automated assessment method previously described. Before these sessions the students took a written pretest, and afterward they took a written posttest. The test was designed to assess students' help seeking in two ways. First, students' ability to take advantage of written hints in the context of problem solving was assessed by comparing student performance on hint and nonhint versions of a number of geometry problems. It was found that the students' ability to take advantage of hints on the written test correlated significantly with the quality of their help-seeking behavior during their work with the tutor, as assessed by means of the model-tracing method. Students who were better at seeking help during their work with the

tutor were also better able to take advantage of embedded hints during the posttest.

As a second measure of help-seeking ability, we included hypothetical help-seeking dilemmas in the posttest, short descriptions of frequently occurring situations during work with the tutor accompanied by a multiple-choice question as to what the appropriate help-seeking behavior would be in the given situation. In a median-split analysis, we found that students who sought help in a more adaptive manner during their work with the tutor (as assessed using the model-tracing method) did significantly better on the hypothetical help-seeking dilemmas. In sum, the model converges with two other measures of help seeking, the ability to take advantage of hints embedded in problem-solving items in a paper-and-pencil test, and the ability to select an appropriate course of action in a hypothetical help-seeking scenario described in words.

EFFECT OF FEEDBACK ON STUDENTS'
HELP-SEEKING BEHAVIOR

A primary motivation for developing the model of help seeking and the assessment method based on the model was to test the hypothesis that feedback on help seeking helps students become more effective help seekers. As a result, students may achieve more robust learning at the domain level. By robust learning we mean learning that lasts, transfers to new situations, or facilitates future learning (Koedinger et al., 2009; Koedinger, Corbett, & Perfetti, 2010). Having embedded the automated method for evaluating students' help-seeking behavior in the Geometry Cognitive Tutor, it was a relatively small step to make the tutor provide feedback on students' help-seeking behavior. We call the resulting tutor the "Help Tutor" (Roll, Aleven, McLaren, & Koedinger, 2007a, 2007b). The Help Tutor assesses each student action with respect to its model of help seeking. When a student action is deemed to be productive help-seeking behavior (i.e., it conforms to the model), the Help Tutor does not provide any explicit feedback related to help seeking. When the student's action is deemed to be undesirable help-seeking behavior the Help Tutor presented a feedback message, as illustrated in Figure 3. These messages are associated with the detailed types of undesirable help-seeking behavior captured in the model (i.e., each error type—implemented as a separate production rule—had its own associated message). The feedback on help seeking was given in addition to the normal guidance that the Geometry Cognitive Tutor provided with respect to the geometry aspects of students' problem-solving activities.

The hypothesis just stated was tested in a study in a vocational school in a rural area outside of Pittsburgh, where the Geometry Cognitive Tutor is in regular use (Roll, Aleven, McLaren, & Koedinger, 2007a, 2007b, in press). It involved 67 students (10th and 11th graders) in four classes and two
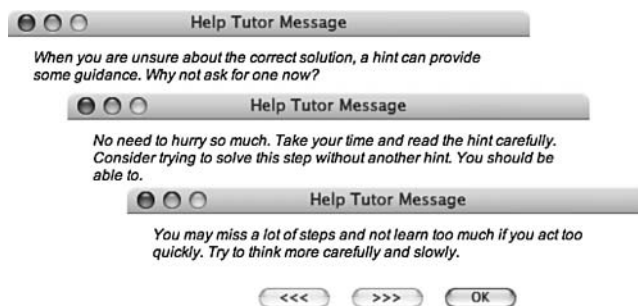

FIGURE 3 Examples of feedback messages from the Help Tutor.

teachers. The study compared the regular Geometry Cognitive Tutor against the Help Tutor, enhanced with two additional elements to support help seeking, which were added to try to amplify the effect of the Help Tutor. Specifically, we added a short video presentation about principles of help seeking, and a small number of newly developed tutor activities in which students self-assessed their ability to solve simple geometry problems. This kind of self-assessment might lead to greater awareness of the need for help, and therefore to more adaptive help seeking (Tobias & Everson, 2002). The experiment covered two 1-month periods, with 1 month in between. During each of these two periods, the instructional activities in each condition covered approximately eight class periods.

The quality of students' help seeking with the tutor was assessed both during and after the intervention. In particular, we extracted from the logs of the student–tutor interactions a number of key variables related to students' use of the tutor and the tutor's help facilities. Consistent with our hypothesis, we saw improvement in students' help-seeking behavior both while the intervention was in effect and after the intervention was turned off. While the intervention was in effect, students who worked with the Help Tutor took more time before deciding to request a hint, spent more time with the hints that they saw, and tended to request fewer hints per hint sequence. The changes in these measures can be viewed as evidence that the students used the tutors' hints in a more deliberate manner, specifically, that the decision to request a hint was made after more deliberation, that they read hints more carefully, and that they needed fewer hint levels in order to make sense out of the hints. The effect persisted after the intervention: The students who had worked with the Help Tutor still spent more time with hints, with an effect size of .9 standard deviations (Cohen's d), and they went less deeply into the tutor's hint sequences, with an effect size of .5 standard deviations (the latter effect was marginally statistically significant). However, these differences in help-seeking behavior did not lead to differences in domain-level learning, which was assessed using four types of items testing students' conceptual and procedural learning, including two types of transfer items (Roll et al., 2007a, in press).

Thus, the Help Tutor, in combination with other forms of instruction on help seeking, led to a lasting improvement in students' help-seeking behaviors. The design of the experiment did not allow us definitively to attribute the effect to the Help Tutor, because the students in the experimental condition also received other instruction related to help seeking, namely, the brief introductory video and self-assessment activities. It seems unlikely, however, that these other components by themselves could have caused the improvement that was observed. The 4-min video presentation was largely ineffective. It was hard to see on the small TV screen that was used, and many students were surfing the web. The self-assessment activities, if anything, should have affected the way students ask for a first hint on a step, but that effect was not observed. The Help Tutor was the only component that was in effect for an extensive amount of time.

We see a number of possible explanations for the fact that there was no improvement in students' domain-level learning. First, it is possible that effective help seeking is not as strong an influence on learning in a tutored problem-solving environment as we assumed. It may well be that learning from tutor hints (which are complex explanations) requires mathematical reading comprehension abilities that the students in the target population often do not possess. Second, it may be that even though students' help use improved quite substantially, still greater improvement is needed before we see a noticeable improvement in domain-level learning. Finally, it may be that the Help Tutor model misses some productive forms of help-seeking behavior and therefore discourages them through its feedback messages. For example, in subsequent research, Shih et al. (2008) found that students who spend more time after receiving a bottom-out hint, but before moving on to the next step in the tutor problem, tend to have better learning results. It appears that these students are self-explaining the bottom-out hints, akin to self-explaining a worked example step. (It would be interesting to confirm this interpretation through think aloud data.) It may be that if the model were extended to account for this behavior, the Help Tutor would be more effective. A more comprehensive model might include other aspects of self-regulation as well, for example, the decision to self-explain when not prompted by the software, the decision to seek help outside the software, and the decision to review a problem solution at the end of a problem. Creating such a comprehensive model is an interesting avenue for future work.

Although we had hoped to find better learning at the domain level, due to feedback on help seeking, it was encouraging to see at least a lasting improvement in students' help-seeking behavior. A lasting improvement in SRL is notoriously difficult to achieve, so the progress made in this work should be viewed positively, just like other instructional programs that have succeeded in this regard (Mevarech & Kramarski, 2003; White & Frederiksen, 1998; Zimmerman & Moylan, 2009).

DISCUSSION

The research reviewed in this article demonstrated a practical method for embedding assessment of students' SRL in CBLEs used in actual classrooms. The method was applied to one particular SRL strategy, help seeking. This method is unobtrusive, meaning that it that does not distract from instructional goals and, ideally, enhances them. It can be applied longitudinally (i.e., over extended periods), at a detailed level both in the temporal sense—multiple observations within minutes—and with respect to a detailed model of desirable and undesirable behaviors. The method successfully addresses a key challenge for assessment methods that view SRL as a sequence of events: It is necessary to interpret observable behaviors depending on the learning context in which they occur. Compared to other methodologies that are increasingly being used to study SRL, such as think aloud protocols (Azevedo et al., 2004) and diaries, the model-tracing approach provides an interesting alternative or complement. In think aloud approaches, for example, aspects of student cognition or self-regulation that are not easily observable in the software such as FOK and JOL are inferred from student self-report. In the model-tracing approach, by contrast these latent variables are inferred from observable actions. At this point we do not know which is the better indicator. On the other hand, analysis of verbal protocol data is notoriously labor-intensive, making this method unsuitable for tracking large numbers of students over extended periods, in contrast to the model-tracing approach. Also, the think aloud method is not easily applied in actual classrooms. Finally, even when interpreting think aloud data, one needs a method for the context-sensitive assessment of student self-regulation: student actions need to be assessed based on the context in which they occur. Similar comments can be made about other observational methods. We therefore concur with Zimmerman's (2008, p. 171) advice that it may be most productive to use automated measures (such as the model-tracing approach discussed in this article) in conjunction with other measures.

As for the generality of our approach to assessing SRL through modeling and model tracing, although the reviewed research applied this approach to one particular SRL strategy, help seeking with an ITS, it could likely be applied to other SRL strategies as well. The main effort involved would be in extending the model to include additional strategies. In a sense, the help-seeking model already captures multiple SRL processes and strategies. As discussed, it models simple forms of FOK and JOL, key SRL processes (Azevedo et al., 2004), namely, in the decision points where students assess whether a step is familiar and whether they have a sense of what to do, and where they decide whether they understand a hint well enough to try the step without additional hint levels (see Figure 2). Further, a few additional models of SRL have been built and applied in tutors, albeit simpler examples than the model reviewed in the current article. Mathan

and Koedinger's (2005) intelligent novice model captures SRL processes, namely, performance monitoring and error self-correction. Also, the self-assessment tutor used in the second classroom study previously reviewed modeled and assessed simple forms of student self-assessment (Roll et al., 2007a). Although these assessment models are not nearly as sophisticated as the Help Tutor, they are examples of generalization Thus, there is evidence that assessing SRL and other learning processes, such as collaborative learning processes (Walker, Walker, Rummel, & Koedinger, 2010) and scientific inquiry (Koedinger, Suthers, & Forbus, 1999; Roll, Aleven, & Koedinger, 2010), is possible using ITS techniques such as model tracing.

The model-tracing method fundamentally reflects a view of SRL as a sequence of events. In particular, the model used to assess SRL reflects this view: It describes recommended help-seeking behavior as sequences of observable and mental actions (summarized in the flow chart of Figure 2). A practical advantage of assessing SRL in an event-based manner is that a system can give fine-grained, action-by-action feedback on students' actual use of SRL strategies, as illustrated in the reviewed work.

From a theoretical perspective, treating SRL as a sequence of events has the advantage that finer grain changes might be observed over time than is feasible with surveys, which for practical reasons cannot be assigned more than a handful of times during any given study. Fine-grained data about SRL can be used by researchers in many ways, for example, to study development over time of students' SRL strategies or to study relations between SRL and domain-level learning in a more fine-grained manner than was previously possible. ITSs have long been capable of producing detailed, event-based data regarding students' learning at the domain level (Corbett et al., 2000; Koedinger, Cunningham, Skogsholm, & Leber, 2008; VanLehn et al., 2007). With the addition of an automated method to assess SRL strategies in a fine-grained event-based manner, an ITS can produce tightly-coupled streams of detailed data, one related to targeted SRL strategies, the other related to domain-level learning. Having these two parallel and interrelated data streams will enable SRL researchers to study, in considerable detail, questions about the relation between SRL and domain-level learning. For example, they could investigate whether students with effective help-seeking patterns early on during the learning of a skill component exhibit more efficient learning of the same skill component later on. As a second example, they could investigate whether learning episodes characterized by effective SRL also seem to be characterized by a relatively fast rate of domain-level learning. More broadly, the analysis of tightly coupled streams of behavioral data at the domain level *and* the self-regulatory level opens up opportunities for tighter links between SRL theories and cognitive theories of learning.

To conclude, the reviewed work focused on automated assessment of a specific SRL strategy, namely, help-seeking

within a tutoring system. Researchers have used the novel assessment method developed in this research to show that this SRL strategy can be improved through a tutoring approach that previously had been demonstrated for domain-specific knowledge. More generally, the approach can be used for temporally fine-grained data collection over extended periods, improving upon panel approaches for purposes of experimental or correlational analysis. The approach to SRL strategy modeling and the model-tracing technology could be applied to other aspects of SRL as well.

ACKNOWLEDGMENTS

REFERENCES

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

Aleven, V., McLaren, B. M., Roll, I., & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16,* 101–128.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R.M. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73,* 277–320.

Anderson, J. R. (1993). *Rules of the mind.* Mahwah, NJ: Erlbaum.

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.

Anderson, J. R., & Lebière, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning, 2,* 57–65.

Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology, 29,* 344–370.

Baker, R. S. J. d. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In M. B. Rosson & D. J. Gilmore (Eds.), *Proceedings of ACM CHI 2007: Computer-Human Interaction* (pp. 1059–1068). New York, NY: ACM.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research, 19,* 185–224.

Bembenutty, H. (2009). Feeling-of-knowing judgment and self-regulation of learning. *Education, 129,* 589–598.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & the Teachable Agents Group at Vanderbilt (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19,* 363–392.

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling & User-Adapted Interaction, 11,* 87–110.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4,* 253–278.

Corbett, A., McLaughlin, M., & Scarpinatto, K. C. (2000). Modeling student knowledge: Cognitive Tutors in high school and college. *User Modeling and User-Adapted Interaction, 10,* 81–108.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68,* 179–201.

Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research, 68,* 322–349.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *The Journal of the Learning Sciences, 8,* 391–450.

Gross, A. E., & McMullen, P. A. (1983). Models of the help-seeking process. In J. D. Fisher, N. Nadler, & B. M. DePaulo (Eds.), *New directions in helping* (Vol. 2, pp. 45–61). New York, NY: Academic Press.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning, 2,* 107–124.

Karabenick, S. A. (Ed.). (1998). *Strategic help seeking: Implications for learning and teaching.* Mahwah, NJ: Erlbaum.

Karabenick, S. A., & Newman, R. (Eds.) (2006). *Help seeking in academic settings: Goals, groups, and contexts.* Mahwah, NJ: Erlbaum.

Koedinger, K. R., & Aleven V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review, 19,* 239–264.

Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 897–964). New York, NY: Routledge.

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*(1), 30–43.

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive Tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). New York, NY: Cambridge University Press.

Koedinger, K. R., Corbett, A. C., & Perfetti, C. (2010). *The Knowledge-Learning-Instruction (KLI) framework: Toward bridging the science-practice chasm to enhance robust student learning* (CMU-HCII Tech Rep. No. 10-102). Available from http://reports-archive.adm.cs.cmu.edu/hcii.html

Koedinger, K., Cunningham, K., Skogsholm A., & Leber, B. (2008, June). An open repository and analysis tools for fine-grained, longitudinal learner data. In R. S. J. d. Baker, T. Barnes, & J. E. Beck (Eds.), *Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings* (pp. 157–166). Montreal, Quebec, Canada.

Koedinger, K. R., Suthers, D. D., & Forbus, K. D. (1999). Component-based construction of a science learning space. *International Journal of Artificial Intelligence in Education, 10,* 166–175.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction, 18,* 495–523.

Lovett, M. C. (1998). Cognitive task analysis in service of intelligent tutoring system design: a case study in statistics. In Goettl, B. P., Halff, H. M., Redfield, C. L., & Shute, V. J. (Eds.), *Proceedings of the 4th International Conference on Intelligent Tutoring Systems, ITS 2008* (pp. 234–243). Berlin, Germany: Springer-Verlag.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist, 40,* 257–265.

Mevarech, Z. R., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology, 73,* 449–471.

Mitrovic, A., & Ohlsson, S. (1999). Evaluation of a constraint-based tutor for a database language. *International Journal of Artificial Intelligence in Education, 10,* 238–256.

Nelson-Le Gall, S. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review, 1,* 224–246.

Nelson-Le Gall, S. (1985). Help-seeking behavior in learning. *Review of Research in Education, 12*(1), 55–90.

Newell, A. & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Newman, R. S. (1998). Adaptive help seeking: A role of social interaction in self-regulated learning. In S. A. Karabenick (Ed.), *Strategic help seeking. Implications for learning and teaching* (pp. 13–37). Mahwah, NJ: Erlbaum.

Newman, R. S. (2008). The motivational role of adaptive help seeking in self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 315–337). New York, NY: Erlbaum.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16,* 385–407.

Roll, I., Aleven, V., & Koedinger, K. R. (2010). The Invention Lab: Using a hybrid of model tracing and constraint-based modeling to offer intelligent support in inquiry environments. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems, ITS 2010* (Vol. 1, pp. 115–124). Berlin: Springer.

Roll, I., Aleven, V., McLaren, B., & Koedinger, K. (2007a). Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education, AIED 2007* (pp. 203–210). Amsterdam, The Netherlands: IOS Press.

Roll, I., Aleven, V., McLaren, B., & Koedinger, K. (2007b). Designing for metacognition – applying Cognitive Tutor principles to metacognitive tutoring. *Metacognition and Learning, 2,* 125–140.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (in press). On-line assessment and tutoring of help seeking: Toward domain learning through metacognitive improvement. *Learning and Instruction.* doi: 10.1016/j.learninstruc.2010.07.004

Roll, I., Aleven, V., McLaren, B. M., Ryu, E., Baker, R. S., & Koedinger, K. R. (2006). The Help Tutor: does metacognitive feedback improve students' help-seeking actions, skills and learning? In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), *Proceedings of 8th International Conference on Intelligent Tutoring Systems, ITS 2006* (pp. 360–369). Berlin, Germany: Springer Verlag.

Schofield, J. W. (1995). *Computers and classroom culture.* New York, NY: Cambridge University Press.

Shih, B., Koedinger, K. R., & Scheines, R. (2008). A response time model for bottom-out hints as worked examples. In R. S. J. d. Baker, T. Barnes, & J. Beck (Eds.), *Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings* (pp. 117–126). Montreal, Quebec, Canada.

Tobias, S., & Everson, H. T. (2002). *Knowing what you know and what you don't: Further research on metacognitive knowledge monitoring* (College Board Research Report No. 2002-3). New York, NY: College Board Publications.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16,* 227–265.

VanLehn, K., Koedinger, K.R., Skogsholm, A., Nwaigwe, A. Hausmann, R. G. M., Weinstein, A. & Billings, B. (2007). What's in a step? Toward

general, abstract representations of tutoring system log data. In C. Conati, K. McCoy, & G. Paliouras (Eds.), *Proceedings of the 11th International Conference on User Modeling, UM 2007* (pp. 455–459). Berlin, Germany: Springer.

VanLehn, K., Lynch, C., Schultz, K., Shapiro, J.A., Shelby, R.H., Taylor, L., . . . Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education, 15,* 147–204.

Walker, E., Walker, S., Rummel, N., & Koedinger, K. R. (2010). Using problem-solving context to assess help quality in computer-mediated peer tutoring. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems, ITS 2010* (Vol. I, pp. 145–155). Berlin, Germany: Springer.

Walonoski, J., & Heffernan, N. T. (2006). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In M. Ikeda, K. D. Ashley & T.-W. Chan (Eds.), *Proceedings of the 8th International Conference on Intelligent Tutoring Systems, ITS 2006* (pp. 382–391). Berlin, Germany: Springer-Verlag.

Wagster, J., Tan, J., Wu, Y., Biswas, G., & Schwartz, D. (2007). Do learning by teaching environments with metacognitive support help students develop better learning behaviors? In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Meeting of the Cognitive Science Society* (pp. 695–700). Austin, TX: Cognitive Science Society.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 279–306). Hillsdale, NJ: Erlbaum.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports and study tactics and achievement. *Contemporary Educational Psychology, 27,* 551–572.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA: Academic Press.

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

Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning.* Burlington, MA: Morgan Kaufmann.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). New York, NY: Routledge.

Zusho, A., Karabenick, S. A., Bonney, C. R., & Sims, B. C. (2007). Contextual determinants of motivation and help seeking in the college classroom. In R. P. Perry, & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 611–659). Dordrecht, The Netherlands: Springer.