

Towards Understanding How to Assess Help-Seeking Behavior Across Cultures

Amy Ogan, *Carnegie Mellon University*

Erin Walker, *Arizona State University*

Ryan Baker, *Columbia University*

Ma. Mercedes T. Rodrigo, *Ateneo de Manila University*

Jose Carlo Soriano, *Ateneo de Manila University*

Maynor Jimenez Castro, *Universidad de Costa Rica*

Abstract. In recent years, there has been increasing interest in automatically assessing help seeking, the process of referring to resources outside of oneself to accomplish a task or solve a problem. Research in the United States has shown that specific help-seeking behaviors led to better learning within intelligent tutoring systems. However, intelligent tutors are used differently by students in different countries, raising the question of whether the same help-seeking behaviors are effective and desirable in different cultural settings. To investigate this question, models connecting help-seeking behaviors with learning were generated from datasets from students in three countries – Costa Rica, the Philippines, and the United States, as well as a combined dataset from all three sites. Each model was tested on data from the other countries. This study found that models of effective help seeking transfer to some degree between the United States and Philippines, but not between those countries and Costa Rica. Differences may be explained by variations in classroom practices between the sites; for example, greater collaboration observed in the Costa Rican site indicates that much help seeking occurred outside of the technology. Findings indicate that greater care should be taken when assuming that the models underlying AIED systems generalize across cultures and contexts.

Keywords: Help Seeking, Cross-Curricular Skills, Assessment, Intelligent Tutoring System

INTRODUCTION

The teaching and assessing of cross-curricular skills that cut across domain boundaries, such as problem-solving and critical thinking, has long been of interest in education (e.g., Judd, 1908). This interest has led to the decision to emphasize cross-curricular skills such as problem-solving in the PISA2012 international examinations, and collaborative problem-solving on the upcoming PISA2015 (Beller, 2011). There has been increasing research on assessing the complex behaviors demonstrated in cross-curricular skills with technology (Martin, 2008), and in modeling them for use in personalized learning technologies (Csapó, 1999; Greiff, 2012). One such skill that has received particular attention in recent years in the AIED community and throughout the learning sciences is how students seek and utilize help.

Help seeking is the process of looking to resources outside of oneself to find information or strategies that will assist in accomplishing a task or solving a problem (Ames & Lau, 1982, Karabenick & Knapp, 1991, Nelson-Le Gall, 1985). Early research regarded help seeking as a negative behavior that indicated student dependency on outside sources (Ames & Lau, 1982, Nelson-Le Gall, 1985). These studies therefore concentrated on help seeking's potential negative effects on the student, such as embarrassment and damage to self-esteem (Nadler & Fisher, 1986, Nelson-Le Gall, 1985). During the eighties, however, the view of help seeking changed (Ames & Lau, 1982, Polson & Richardson, 1988). Educational researchers recognized that it was important for students to seek help, especially when they encountered ambiguity or difficulty in schoolwork, in order to be able to continue with the learning process (Ames & Lau, 1982).

Here, we focus on the role of help seeking within online learning for multiple reasons. First, help seeking that occurs online is an increasingly important part of learning, as growing amounts of

learning take place online around the world (Nafukho, 2007; O'Lawrence, 2005). Second, online help seeking is likely to be easier to assess behaviorally than offline help seeking, especially given the recent advances in online methods for assessing help seeking (cf. Alevan et al., 2006). This increased ease of assessment becomes a particularly important consideration for conducting assessment in a cross-cultural context, where it may be difficult to directly observe the help seeking that is occurring in classrooms. However, as we describe in more detail below, few studies have investigated our ability to assess help seeking skills across cultures.

In this paper, we therefore consider how the cross-curricular skill of help seeking differs across cultures. Behaviors for many cross-curricular skills may differ between cultures; for instance, collaborative behaviors that are considered desirable in one culture may be considered problematic or even offensive in other cultures (Kim & Bonk, 2002; Vatrappu, 2008). It has been noted that most of the research into student interactions within educational software – such as research on help seeking – takes place in wealthy North American countries, European countries, and Australia/New Zealand (Blanchard, 2012). As such, relatively little is known about how cultural differences might affect key features of help seeking with learning technologies. There is a need for greater research on how computer-based assessments of specific skills can account for cultural norms.

Objectives and Research Questions

The main objective of this study is to investigate whether the assessment of effective help-seeking behavior can be consistent across cultures. To do this, we will use data mining methods to develop models that predict effective learning from students' help-seeking behaviors, using data sets from sites in three countries that are expected to vary along a number of cultural dimensions: Costa Rica, the Philippines, and the U.S.A. We will then compare between the models found for the three sites to see how effective help-seeking behaviors are similar or different across cultures. While culture is notoriously difficult to define, we work with the understanding that it entails “a fuzzy set of basic assumptions and values, orientations to life, beliefs, policies, procedures and behavioral conventions that are shared by a group of people, and that influence (but do not determine) each member's behavior” (Spencer-Oatey, 2008). In this paper, we link it operationally with the three countries being studied, under the viewpoint that the cultural differences between these countries are broadly greater than the cultural differences within them, and that many – though not all – of the differences between these three countries can be understood in terms of culture.

This study aims to answer the following research questions:

1. Which help-seeking behaviors are effective in online learning within sites located in Costa Rica, the Philippines, and the United States of America?
2. How well does a model of effective help-seeking behavior generalize across three cultures?

By understanding the cross-cultural variations in this essential cross-curricular skill, we can understand whether and how this skill should be assessed differently in different cultural contexts, and work towards achieving a less culturally biased picture of this skill's application in different societies.

Effective vs. Ineffective Help Seeking

Help seeking may be characterized as effective or ineffective. Effective help seeking takes place when a learner knows when he needs help, what kind of help he needs, whom to ask for help, and how to ask for the help that he needs. Effective help seeking may avert possible failure, maintain engagement, and lead to long-term mastery and autonomous learning (Newman, 2002). Effective help seeking is viewed as an important strategy that contributes to self-regulation and vice versa (Newman, 2002, Puustinen, 1998, Ryan, Hicks, & Midgley, 1997), enabling the student to continue learning. Self-regulated students control the frequency with which they ask for help, asking only at appropriate times and avoiding dependence (Puustinen, 1998). Both the overuse of help and the avoidance of help when it is needed (Ryan, Pintrich, & Midgley, 2001) may lead to less effective learning.

Newman (2002) identifies four specific competencies and motivational resources for effective and adaptive help seeking:

- (a) cognitive competencies (i.e., knowing when help is necessary, knowing that others can help, knowing how to ask a question that yields precisely what is needed);
- (b) social competencies (i.e., knowing who is the best person to approach for help, knowing how to carry out a request for help in a socially appropriate way);
- (c) personal motivational resources (i.e., personal goals, self-beliefs, and feelings associated with tolerance for task difficulty; willingness to express to others a need for help; and a sense of personal agency);
- (d) contextual motivational resources (i.e., classroom factors such as goals, grading systems, collaborative activities, student-teacher interaction, and teacher expectations for the child that facilitate help seeking).

Beyond simply whether a student seeks help when it is needed, it is important to consider the degree to which the student thinks about the help received (cf. Bielaczyc, Pirolli, & Brown, 1995), an act termed self-explanation. Self-explanation behaviors have been found to be associated with positive learning outcomes across different types of learning contexts and with learning from different types of help (Baker, Gowda, & Corbett, 2011; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Shih, Koedinger, & Scheines, 2008; VanLehn, Jones, & Chi, 1992).

By contrast, some students abuse help rather than thinking through the help and trying to learn from it. Help abuse refers to the use of help only to advance in the curriculum or obtain the solution to a particular problem, rather than thinking through the subject matter (Alevin et al., 2006); this behavior has also been termed executive help seeking (Nelson-Le Gall, 1985) and gaming the system (Baker, Corbett, Koedinger, & Wagner, 2004). This behavior may prevent students from learning because it bypasses the self-explanation and self-regulation processes.

Help Seeking in the Traditional Classroom

In the past few decades, there has been considerable research within traditional classrooms to understand the differences in students' help-seeking behaviors. A variety of methods have been employed, including questionnaires (Karabenick & Knapp, 1991; Ryan et al., 1997; Taplin, Yum, Jegede, Fan, & Chan, 2001), analysis of student help-seeking choices during learning (Ames & Lau, 1982), and interviews (Taplin, et al., 2001). Across this body of research, a considerable amount has been learned about who seeks help; for instance, Ames & Lau (1982) find that student beliefs about the usefulness of help sessions influence the choice to attend help sessions. Similarly, Hofer et al. (1996) (cited in Alevin et al., 2003) identified more positive opinions of help when it was perceived as relevant to the task being performed. Karabenick & Knapp (1991) found evidence that learners who seek more active control of their learning are more likely to ask for help. However, conflicting reports have been made about whether generally more successful or generally less successful students are likely to seek help (Ames & Lau, 1982; Karabenick & Knapp, 1991; Tapin et al., 2001).

Help Seeking Within Online Learning

In recent years, the findings derived from studies in traditional classrooms have been augmented by research on student help seeking within online learning. Unlike research based on questionnaires and interviews, online learning environments can provide rich traces of every action a student makes, allowing a researcher to study not only what a student does, but the context of the action in terms of the student's recent experiences, and the impacts of the student's decision over the next minutes. While analysis of student help-seeking behavior occurred before the use of data from online learning environments (cf. Ames & Lau, 1982), it was very time-consuming and difficult to analyze behavior at a fine-grained scale over large numbers of students. By contrast, online learning environments make it easy to collect fine-grained data from hundreds or thousands of students, and data mining methods make analysis very scalable. In addition, learning systems such as intelligent tutoring systems (ITS; Koedinger & Corbett, 2006; Woolf, 2009) often provide students with help of different types and modalities (Woolf, 2009) including hints, glossary support, scaffolds that break problems down, and

directed messages about student misconceptions. Some ITS provide different levels of help- from general guidance to bottom-out hints (Anderson, Boyle, & Reiser, 1985; Wood & Wood, 1999). In addition, help is sometimes given automatically and sometimes given upon request by the student, allowing a researcher to analyze a student's differential responses to these two situations. This diversity of types and contexts of help available makes more sophisticated comparisons of different types of help seeking very feasible.

There has been an increasing amount of research within this paradigm in recent years. In one of the earliest studies on help seeking within online learning, Wood and Wood (1999) found that students with more prior knowledge tended to work faster, make fewer errors, and ask for less help, and that students who requested less help actually achieved better learning outcomes. A contrasting pattern was found by Alevan and colleagues (2004), who found a positive relationship between help seeking and learning. Baker, Gowda, & Corbett (2011) similarly found that the failure to seek help was correlated with a lower degree of student preparation for future learning.

However, the way the student uses the help they seek matters. For example, students who pause to think through the implications of a hint have higher learning gains (Shih et al., 2008), and time spent on problems where help was sought is positively related to learning gains (Arroyo and Woolf, 2005). In contrast, intentional misuse of help features to obtain answers without thinking, sometimes by clicking through hints rapidly, is associated with poorer learning (Alevan et al., 2004; Baker et al., 2004). It also matters which type of help a student uses; students who most frequently use low-level help (which is more explicit) tend to show lower learning (Mathews, Mitrovic, & Thomson, 2008) than students who focus on help related to concepts.

Help Seeking Across Cultures

One limitation of much of the research in help seeking in online learning environments is that it has still largely taken place in fairly similar and homogenous populations. Indeed, most of the data in the studies listed above was drawn exclusively from one region of the United States (Wood & Wood were using data from Britain; Mathews, Mitrovic, & Thompson were using data from New Zealand). It is unclear to what degree these relationships between help seeking and learning are consistent across different populations and cultures. It is logical to assume that the parameters of Newman's help seeking competencies may be affected by the culture in which they are expressed – for example, the most appropriate source of help to approach in one culture may be the expert (e.g., the teacher), while in another, the expert should not be bothered until other sources are exhausted.

While there has only been a limited degree of research comparing help-seeking behaviors across cultures, online or offline, some noteworthy examples exist that would support this claim. For instance, a study comparing help-seeking orientations of Latinos and non-Latinos in a public urban high school in the United States, found that non-Latinos and English-speaking Latinos exhibited the same desire for academic support, while Spanish-dominant Latinos exhibited a significantly lower desire for academic support (Stanton-Salazar, Chavez, & Tai, 2001). The researchers conjectured that this difference might be the result of the influence of American culture on the English-proficient Latinos. Another comparison of learners in the United States and Oman found that Americans reported less self-regulation of help seeking than Omani students (Al-Harathi, 2008). In a recent study by Ogan, Walker, Baker, Rebolledo, & Jimenez-Castro (2012) within the context of classrooms using educational software, researchers found that students in Mexico, Costa Rica, and Brazil using an intelligent tutor tended to work far more collaboratively than their American counterparts. Software that was intended for individual use became the object of group activity, and the primary source of help in these cases was not the educational software but other classmates.

It is still not known, however, whether a) help-seeking skills can be modeled in a generalizable way, or b) despite differences in the cultural appropriateness of help-seeking behaviors, a standard set of help seeking approaches lead to greater learning.

METHOD

As previously discussed, the main objective of this study is to investigate whether effective help-seeking behavior is similar or different across cultures, towards understanding how assessment of this cross-

curricular competency might be similar or different across cultures. In this section, we describe how we created machine-learned models for predicting learning through help-seeking behavior from datasets from sites in three countries: Costa Rica, Philippines, and U.S.A. The data sets consisted of log files recording student interactions with a Cognitive Tutor for teaching generation and interpretation of scatterplots. The data from the three sites has also been used in previous studies, but the help-seeking behaviors of the students have not previously been analyzed.

Data Source

This study had two main sources of data: log files from student interaction with a Cognitive Tutor for Scatterplot generation and interpretation (Baker et al., 2006), and pre-test and post-test scores that assessed student knowledge before and after interaction with the system. The data was collected as part of a classroom-based study in which participating mathematics teachers were first provided training with the system and then asked to conduct each session as if they were including the technology as part of their typical practice. In each case, class was held in the computer lab during the regularly scheduled math period. Researchers were available for technical support and to ensure that the study procedure was kept as consistent as possible across the settings. Additionally, at least two researchers were present in the computer lab in each session, taking field notes as they positioned themselves around the lab in order to observe computer screens. Field notes captured on- and off-task behavior including collaboration, teachers' instructional procedures, student impasses, and affective reactions.

In all three sites, the Scatterplot Tutor was used for a total of 80 minutes, after brief instruction on the domain concepts and a pre-test. In the tutor, students created scatterplots of data for small data sets, and then interpreted the data sets; the process of scatterplot creation, from selecting appropriate variables, to choosing axis values, to plotting points was reified (e.g. each cognitive step was made visible), and computer-generated hints were available at each step. In addition, bug messages were given when student actions indicated a known misconception, such as the belief that all graphs should have one categorical variable and one numerical variable. Within this Cognitive Tutor, an animated agent was incorporated named "Scooter the Tutor", developed using graphics from the Microsoft Office Assistant (Microsoft, 1997) but modifying those graphics to enable a wider range of emotions. Scooter was designed to reduce the incentive to game the system, through expressions of negative emotion, and to help students learn the material that they were avoiding by gaming, by providing supplementary exercises after gaming episodes. Scooter was found to reduce gaming behavior in the U.S.A and improve gaming students' learning (Baker et al., 2006), but was not found to be effective at reducing gaming in the Philippines (Rodrigo et al., 2012).

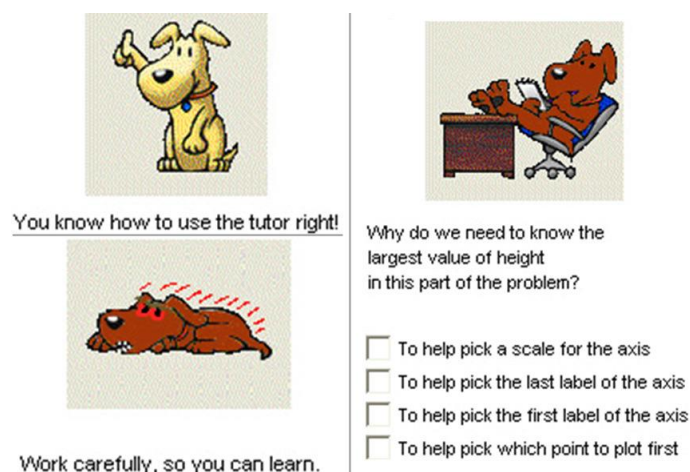


Fig. 1. Scooter the Tutor – looking happy when the student does not game (top-left), giving a supplementary exercise to a gaming student (right), and angry when the student is believed to have been gaming heavily, or attempted to game Scooter during a supplementary exercise (bottom-left).

A post-test was given in class after the final tutor session. The pretest and posttest consisted of two isomorphic tests counterbalanced across students. In each test, the student had to generate a single scatterplot using provided data and was assessed on choosing appropriate variables, creating appropriate scales, labeling axes, and correctly placing points on the plot (cf. Baker et al., 2006).

Efforts were made to choose comparable schools in the three countries. There is no easy matching of schools between three such different countries, as the overall structure of how cities and metro regions are organized and populated differs greatly between countries. Great efforts were instead made to ensure that the three populations represented comparable demographics within their respective countries. The participating students in all three sites were from the local ethnic majority and were drawn from a public school population that was neither unusually wealthy nor experiencing unusual degrees of poverty relative to the local societal context. The populations in all three sites had an approximately equal number of male and female students, who were from the same year in school yet ranged in age from 12 to 14 years. Given the inherent difficulty in such international comparisons, Table 1 is provided to summarize several important points of comparison that can contextualize the present results for future discussion. In Costa Rica, 85 students participated, in the United States 139 students, and in the Philippines, 127 students.

Table 1. Comparison of the three populations on math skills, rank on the 2012 Program for International Student Assessment math test, computer skills and use, city size, and 2008 gross domestic product per capita in the city of our field site.

	U.S. Site	Philippines Site	Costa Rica Site
# Participants	139	127	85
Age of Participants	12-14	12-14	12-14
Math Skills	learning algebra, basic geometry	learning algebra, linear equations	learning trigonometry, mastered basic math
Computer Lab	1 computer lab, used in writing classes	2 computer labs, used in computer literacy classes	1 computer lab, used across classes
City Population	17,000	2.76 million	10,000
City GDP Per Capita	\$41,500 U.S.D	\$2,760 U.S.D	\$6,590 U.S.D
2012 PISA Score (Avg: 494)	481	does not participate	407

Feature Selection

From log files of student actions within the tutoring software, seventeen aspects of student behavior were distilled. While inspired by extensive prior theory regarding effective and ineffective help-seeking behavior, specific features were directly drawn from recent work to study student preparation for future learning in a similar intelligent tutoring system (e.g. Baker, Gowda, & Corbett, 2011).

All features were normalized as a percentage of total actions taken by the student in the interface. The features were as follows, in Table 2:

Table 2

List of 17 help-seeking features extracted from log data, along with whether they indicate positive or negative help-seeking behaviors.

Code	Feature Name	Feature Description	Predicted Relationship to Help Seeking
NHP	<i>Help avoidance</i>	Does not request help on poorly known skills (Alevén et al., 2006)	Negative
HP	<i>Help use on poorly known skills</i>	Requests help on poorly known skills (cf. Alevén et al., 2006)	Positive
HPQ	<i>Help Abuse</i>	Requests help on poorly known skill but answers quickly afterwards (cf. Alevén et al., 2006, Baker et al., 2004); subset of feature HP.	Negative
HPL	<i>Self-explanation after help use on poorly known skill</i>	Requests help on a poorly known skill and makes a long pause; subset of feature HP.	Positive
NHW	<i>Help non-use on well-known skills</i>	Does not request help on an already known skill (cf. Alevén et al., 2006)	Positive
HW	<i>Help use on well-known skills</i>	Requests help even when skill is already known	Negative
STHW	<i>Short pause after help use on well-known skill</i>	Requests help even when skill is already known, but takes only a short time before the next action; subset of feature HW.	Negative
LTHW	<i>Long pause after help use on well-known skill</i>	Requests help even when a skill is already known, but takes a long pause; subset of feature HW.	Positive
LTB	<i>Long pauses after bug message</i>	Takes a long pause after receiving a bug message (Baker, Gowda, & Corbett, 2011)	Positive
STB	<i>Short pauses after bug message</i>	Does not take much time before the next action when receiving a bug message (Baker, Gowda, & Corbett, 2011)	Negative
LTAH	<i>Long pauses after requesting hints</i>	Takes a long pause after receiving a hint message (Baker, Gowda, & Corbett, 2011)	Positive
STAH	<i>Short pauses after requesting hints</i>	Does not take much time before the next action when receiving a hint message (Baker, Gowda, & Corbett, 2011)	Negative
LTHR	<i>Long pauses after requesting hint and getting current action right</i>	Takes a long pause after receiving a hint message and gives a correct response for the problem step (Shih, Koedinger, & Scheines, 2008).	Positive
STHR	<i>Short pauses after requesting hint and getting current action right</i>	Does not take much time before the next action when receiving a hint message, but gives a correct response to the problem step.	Negative
STBH	<i>Short pause before requesting hints</i>	Spends little time on the action prior to asking for help.	Negative
LTBH	<i>Long pause before requesting hints</i>	Spends a longer interval on the action prior to asking for help (Alevén et al., 2006, Wood & Wood, 1999)	Positive
HF	<i>Help use on first encounter of skill</i>	Requests a hint on the first time student encounters a skill.	Ambiguous

The features were chosen in order to encompass a wide array of possible help-seeking behaviors for students interacting with the Scatterplot tutor, and cognitive tutors more generally. Feature NHP (help avoidance) has been identified in literature as ineffective help-seeking behavior (Ryan, Pintrich, & Midgley, 2001), while its converse, Feature HP (help use on poorly known skills) is identified as both common and desirable. Features HPQ and HPL examine whether students pause after seeking help on an unknown skill. Feature HPQ may indicate hint abuse/executive help seeking/gaming the system. Feature NHW (not using help when it is not needed) is thought to be effective and desirable help-seeking behavior because it is indicative of self-regulation (Newman, 2002), while the converse, Feature HW (using help when it is not needed) may be an indicator of unnecessary student dependence on help.

Features STHW and LTHW examine student pauses when requesting help on a skill the student already knows. Feature LTHW may suggest that the student is continuing to self-explain, even after being able to perform effectively; this could potentially imply that the student does not know the skill as well as the system believes, or that the student is continuing to consider some aspect of the skill. It could also mean that the student did not realize the skill applied in this situation. Feature STHW instead may indicate “gaming the system” on already-known material, perhaps because it is time-consuming (cf. Baker, Corbett, & Koedinger, 2004).

Features LTB and STB involve bug messages (messages explaining why a student made an error), which can be considered a form of help automatically provided by the tutor; these features investigate whether students pause to read and self-explain these messages. It is important to note that the evidence of self-explanation in these features is indirect, but prior work has found a relationship between these pauses and learning (cf. Baker, Gowda, & Corbett, 2011). Features LTAH, STAH, LTHR, and STHR examine whether students pause to self-explain after asking for help (in the case of features LTHR and STHR, after requesting help and entering the correct answer). Again, the evidence for self-explanation is indirect, but prior work has found a relationship between these pauses and learning (cf. Shih et al., 2008).

Features STBH and LTBH examine how long students attempt a problem prior to requesting help from the system. Taking longer before requesting help may be indicative of the student trying to understand the problems carefully before asking for help. Finally, feature HF, hint requests the first time the student encounters a new skill, may either indicate help abuse or a general desire to understand a problem better before attempting it.

Given this list of features, we can infer which help-seeking behaviors are effective and ineffective in each context, by attempting to predict student learning, operationalized as the student’s post-test score minus their pre-test score, showing how much learning students gained. We determined the proportion of student actions that belong to each feature category, and then create combined models to predict effective help seeking for each site. Our process for creating models of effective help seeking for each culture involved several steps: feature engineering (discussed immediately above), feature selection, feature optimization, model creation, and model evaluation.

Feature Optimization

The first step to using the data features discussed above is to select optimal feature values, e.g. the value for each feature parameter that best predicts learning (post-test minus pre-test). For example, feature 1 refers to poorly known skills, but it is necessary to operationalize “poorly known” in terms of a percentage probability that the student knows the skill. All features except for feature 17 incorporate parameters, either time or knowledge (or both).

Knowledge is operationalized within these features using Bayesian Knowledge Tracing (Corbett & Anderson, 1995), the most widely used knowledge assessment algorithm within intelligent tutoring systems. Bayesian Knowledge Tracing uses a Bayesian update process to repeatedly re-estimate recent and current student knowledge based on student correctness and general probabilities of learning, guessing (obtaining correct answers by luck), and slipping (obtaining incorrect answers despite knowing the skill), computed across all students in each country. Bayesian Knowledge Tracing produces an estimate of the probability that a student knows a skill at a given point in the learning sequence.

For feature optimization, we used the method in Baker, Gowda, & Corbett (2011). For each dataset, we performed a brute-force grid search on all features to find the optimal threshold for each feature.

The features were distilled from the logs depending on the thresholds in the grid. For thresholds based on estimates of student knowledge, the grid size for the search was 0.05 (e.g. increments of 5% probability). For thresholds requiring a time interval, the grid size was 0.5 seconds. For each feature and threshold, we performed a single-parameter linear regression with leave-one-out-cross-validation (LOOCV; Efron & Gong, 1983), where a model is repeatedly built on all students but one, and tested on the held-out student; leave-one-out-cross-validation is mathematically equivalent to the Akaike Information Criterion (AIC; Akaike, 1974). Information criteria and cross-validation approaches are increasingly seen as alternatives to the classical statistical significance paradigm, as they assess how well the model will function on new data drawn from the same distribution as the original distribution (e.g. new students from the same population), rather than just the probability that the results seen would have arisen if only chance events were occurring (cf. Raftery, 1995). The cross-validated correlation (computing the correlation between the model's predicted learning and the student's actual learning, within new students) is used as the model goodness criterion. The threshold that showed the highest positive cross-validated correlation for each feature became the threshold that was used for that feature during model creation. As a control against over-fitting, features with optimal thresholds that had negative cross-validated correlation (e.g. the relationship flips in direction between students) were not considered during the creation of the full model, as negative cross-validated correlation means the model fails to generalize to new data.

Model Creation and Evaluation

For each site's data set, we then created a model using forward selection (Ramsey & Schafer, 1993), finding the model that best predicts each student's degree of learning (post-test minus pre-test). The feature whose single-parameter linear regression model had the highest cross-validated correlation was added to the model. Then, the feature that most increased the cross-validated correlation was repeatedly added to the model, one by one, until no additional feature improved the cross-validated correlation.

A fourth model was created by combining the three individual data sets, representing a "multinational" model of effective help seeking, in order to test whether it is possible to have one effective help-seeking model trained from different sets of cultures that generalizes effectively to individual cultures.

As with the single-feature models used during feature optimization, model goodness was assessed using cross-validation, to derive an assessment of likely model performance on new students drawn from each population. The four models were then applied to all data sets, with non-cross-validated correlations used as the measure of goodness.

RESULTS AND ANALYSIS

In this section, we present the final models generated through feature selection for each site. Each site's model is then applied to each of the four data sets, to study the relationship between effective help-seeking behaviors in the different cultures.

Learning Gains

Pre and post test scores for each site are shown in Figure 2. A repeated-measures ANOVA with test time as a within-subjects factor and country as a between subjects factor shows that students in all three sites had significant learning gains from pre-test to post-test ($F(1,348) = 236.81$; $p < 0.001$), and that there is a significant country by test time interaction ($F(2, 348) = 23.51$; $p < 0.001$). A one-way ANOVA shows that there are significant differences between groups at pre-test ($F(2,348) = 65.06$; $p < 0.001$). Tukey post-hoc comparisons of the three groups indicate that the U.S. site ($M = 0.634$, $SD = 0.391$) had higher prior knowledge than both other groups ($p < 0.001$), while the Costa Rica site ($M = 0.331$, $SD = 0.297$) had higher prior knowledge than the Philippines site ($M = 0.181$, $SD = 0.271$), ($p < 0.001$). Hence, some of the greater gains seen in Philippines and Costa Rica are due to students in those countries catching up to students in the U.S.A.

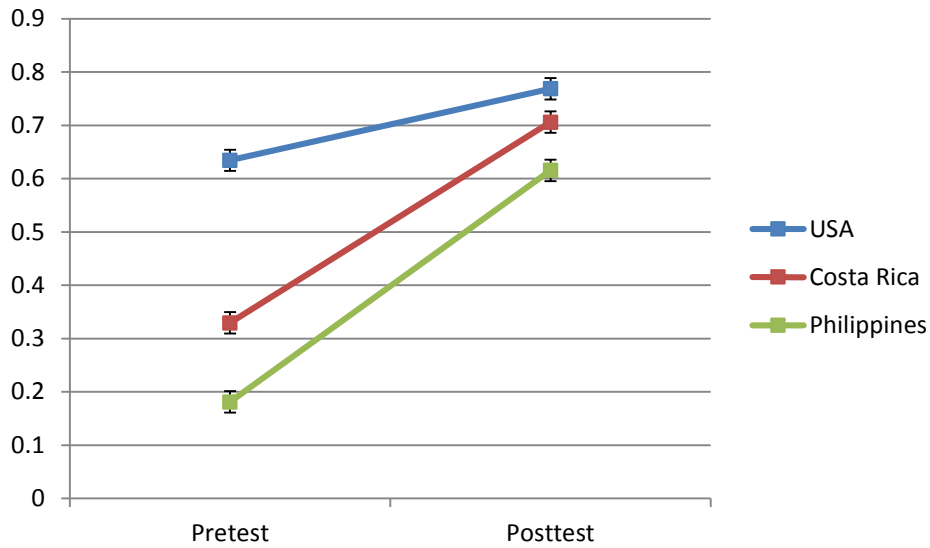


Figure 2. Pre and posttest scores for each site

Forward Selection Results

At a surface level, there were few similarities between the features in the final model for each country, as shown in Table 3. Only two features appeared in the final model of more than one data set: *help avoidance* and *help abuse*. Interestingly, these two features – among the 17 features used in the data set – were drawn from some of the earliest work in modeling help seeking within the log files from educational software (cf. Aleven et al., 2004). *Help abuse (HPQ)* is negatively associated with learning in both the Philippines and United States data sets (in the case of the United States data set, this is a direct replication of Aleven et al., 2004). By contrast, *Help avoidance (NHP)* has a negative feature coefficient for Costa Rica and a positive coefficient for the combined data set. This is not just a case of statistical suppression or lack of attention to collinearity in a model with too many parameters – the relationships go in the same direction in single-feature linear regression models. This suggests that help avoidance is associated with positive outcomes in the U.S. and Philippines data sets, but leads to negative outcomes in the Costa Rica data set.

Table 3

Models of how help-seeking behaviors influence learning in each site. Refer to Table 2 for feature definitions. Thresholds for each parameter are written in parentheses: PK indicates probability that student knows the skill, T indicates time in seconds

Site	Learning =	Cross-validated r
Costa Rica	- 0.132 * NHP (PK < 0.15) + 7.385 * LTAH (T > 47.5) - 9.096 * LTHR (T > 41.5) - 21.847 * HPL (PK < 0.25, T > 45.5) + 53.010	0.462
Philippines	- 0.763 * HPQ (PK < 0.4, T < 1) + 0.021 * NHW (PK > 0.95) + 32.423	0.126

U.S.A	- 6.680	* HPQ	(PK < 0.25, T < 19.5)	0.350
	- 1.021	* HP	(PK < 0.25)	
	- 2.870	* LTB	(T > 57)	
	+ 5.605	* LTBH	(T > 37.5)	
	+ 12.086			
Combined	+ 0.046	* NHP	(PK < 0.05)	0.216
	- 0.482	* SPB	(T < 3.5)	
	- 0.446	* HPL	(PK < 0.35, time > 0.5)	
	+ 8.147	* LTHW	(PK > 0.95, T > 58.5)	
	+ 35.067			

Within the Costa Rica data set, several features were predictive of learning: help avoidance (NHP), long pauses after requesting hints (LTAH), long pauses after requesting hint and getting correct action right (LTHR), and self-explanation after help use on poorly known skill (HPL). The overall cross-validated correlation of the Costa Rica model to the Costa Rica data was a relatively high 0.462.

Within the Philippines data set, only two features were predictive of learning: help non-use on well-known skills (NHW), and help abuse (HPQ). No features overlapped directly or conceptually with the Costa Rica model. The overall cross-validated correlation of the Philippines model to the Philippines data was a fairly low 0.126.

Within the U.S.A data set, several features were predictive of learning: help use on poorly known skills (HP), long pauses after receiving bug messages (LTB), help abuse (HPQ), and long pauses before requesting hints (LTBH). It is worth noting that help abuse was found in both the U.S.A and Philippines models, with help abuse being associated with poorer learning in both data sets. The overall cross-validated correlation of the U.S.A model to the U.S.A model was 0.350.

The multi-national model fit on all 3 data sets performed fairly poorly, achieving a cross-validated correlation of 0.216. We re-visit this model below, in studying its performance on each of the national populations.

Cross-Cultural Evaluation of Models

When each site's model was applied to each of the other data sets, all of the models performed less well (see Table 4), except for the U.S.A model used on the data from the Philippines. However, only two models returned negative correlations to student learning in the new data sets (indicating that the relationships captured point in the same direction in both contexts). Note that these correlations are not cross-validated, as we are evaluating the previously constructed models against the full data sets drawn from each site.

Table 4.

Model Evaluation (non-cross-validated correlation between predicted learning from help-seeking behaviors, and actual learning). Rows are the models, columns are the data sets applied to.

<i>Site</i>	Costa Rica	Philippines	U.S.A	Combined
Costa Rica	0.534	0.051	0.151	0.099
Philippines	0.004	0.203	0.146	0.119
U.S.A	-0.085	0.228	0.476	0.081
Combined	-0.073	0.091	0.238	0.286

In ten out of the 12 cases where a model is applied to the full data set from a site, the correlation between the model's predicted learning and the actual learning was positive, suggesting successful transfer of these models. A sign test can be applied in this case, and indicates two-tailed $p=0.04$; hence, in general the models transfer more than would be expected solely due to chance.

Accounting for Pretest

Given the differences in pretest scores between groups that were observed earlier, we investigated whether the variance we see in the models is related to prior knowledge rather than culture. One way to answer this question is to predict posttest scores, using pretest score as an additional factor, rather than predicting pre-post gains.

When we ran the same models with these new parameters, pretest scores passed the brute-force grid search for all data sets/sites. However, pretest was not selected by forward selection as a predictor of posttest scores in any of the four data sets, and student behaviors did not disappear from the models when pretest scores were included as possible predictors. It is also worth noting that the two sites with the most significant differences in pretest score were the most similar in predictive help-seeking features. As such, student behaviors seem to be better than pretest scores at predicting posttest scores, suggesting that the differences seen between cultures in the help-seeking behaviors associated with learning were not mainly due to pretest differences.

DISCUSSION

We observed distinct differences in the performance of the models across countries. The extremely poor performance of the U.S. model on the Costa Rica data set implied that effective help-seeking behaviors are distinctly different between the U.S. and Costa Rican sites, corresponding to previous findings of differences in how students use educational software in the U.S. and Costa Rica (Ogan et al., 2012, Walker et al., 2011). Overall, Ogan et al. (2012) and Walker et al. (2011) found that the Costa Rican students were more likely to work collaboratively over extended periods of time, even abandoning their own computers to work on a problem together before returning to enter responses, than the American students. The collaborative behaviors seen in Costa Rican students in previous research may therefore explain the difference in help-seeking behavior seen here, as a more collaborative environment may make other students the main source of help while using educational software. As such, the types of help sought from the software and the situations within which it is sought may differ considerably between the two sites.

By contrast, the U.S. model performs even better on the Philippines data than the model built on Philippines data does. This successful transfer suggests that many of the same help-seeking behaviors are associated with effective learning in the two sites. (The Philippines model also performs about as well in the U.S. as it does in the Philippines). Baker et al. (2004) has previously shown that American classrooms using the same system spend only 4% of their time talking to other students or the teacher while working; informally, observations of the students in our Philippines site suggested a similar proportion of collaborative time, making the system the primary source of help. Interestingly, another automated detector has been shown to generalize between students in the U.S. and Philippines; San Pedro, Baker, and Rodrigo (2011) found that a model of carelessness built on data from the U.S. functioned effectively within data from the Philippines, and vice-versa. In contrast, Rodrigo, Baker, and Rossi's (in press) findings, found disengaged behavior to be different between U.S. and the Philippines (in brief, the U.S. students went off-task much more frequently, while students in the Philippines gamed the system more). Together, these findings suggest that even when the appropriateness and efficacy of some learning practices are consistent across cultures, they may not all be.

Even if the help-seeking behaviors are in some part distinct across these sites, we might hypothesize that a combined model that uses all of the data could still be useful for developing personalized learning systems. We found that the model built for the combined multi-national data set performed particularly well on the U.S. data set, moderately on the Philippines data set, and not very well on the Costa Rica data set. If the U.S. and the Philippines are relatively similar, it makes sense that they would dominate prediction in a combined data set. In fact, no model except for the model trained on Costa Rica data did

well on the Costa Rica data set. This suggests that effective help-seeking behavior in Costa Rica is quite different from the other two countries, making combined models difficult. We again note that past research has indicated that Costa Rican students collaborate in more extended fashions when using educational software (Walker et al., 2011; Ogan et al., 2012). Interestingly, several features in the Costa Rican model show that requesting help and pausing are associated with positive learning, a difference from other models. It may be that these long pauses actually represent students seeking help from the software, and discussing the help with other students. Understanding this difference better will be a valuable area of future work.

It is also worth noting that the Costa Rica and U.S. models performed relatively better than the Philippines and combined data sets. It would be possible to conjecture that the poorer performance of the combined data set is caused by the mixed influences of the three different cultures. The diverse mix of help-seeking behaviors makes it more difficult to correctly predict learning for a given data set. In contrast, it is possible that Costa Rica and U.S. students' behavior is more homogenous, making their learning performance more predictable according to observable help-seeking behavior. The Philippines' cross-validated r was the lowest among the four cultures. Analyzing the model, it has the lowest number of features surviving the forward selection process. The implications of this could be 1) Student help-seeking behavior in the Philippines may simply be not as correlated to learning compared to other cultures, 2) Help-seeking behavior in the Philippines may not be as homogenous as in Costa Rica and U.S., and distinct help-seeking strategies may be adopted by students which lead to different patterns of behavior being effective for different sub-groups of students, or 3) The initial feature set was generated based on research that focuses on data from the U.S., and there exist other indicators in the Philippines data of successful or unsuccessful help seeking that have not been captured in these features.

Interpretation Within a Cultural Framework

It is possible to speculate on interpretations of these behavioral findings that stem from differing underlying cultural values. In order to do so, an explanatory framework for culture should be proposed. As an example, Hofstede's dimensions is one of the most frequently cited cultural frameworks given the breadth and extent of his investigations across nations (e.g., Hofstede et al., 2010). One of these dimensions that is particularly salient from this framework for explaining educational results is labeled Masculinity, or "the belief that the society will be driven by competition, achievement and success, with success being defined by the 'winner' or 'best-in-the-field'". Respondents to Hofstede's cultural dimensions survey in the Philippines and the United States indicated that they placed a relatively strong value on a Masculine society, while Costa Rican respondents were quite low on this dimension. Perhaps students in the Philippines and United States viewed work on the ITS as a competition in which getting further ahead was preferable to helping their classmates succeed, while students in Costa Rica may have instead viewed the work as an opportunity to "care for one another" rather than "stand out among the crowd" (a Masculine expression of values).

An alternative interpretation that uses the same framework might be related to the dimension of Individualism, given the observations that the students in the United States and the Philippines for the most part worked individually rather than relying on peers as part of the support. Individualism refers to "the degree of interdependence a society maintains among its members", with a very individualist society placing greater value on autonomy than on group membership. Indeed, Hofstede's survey demonstrated that on average, Costa Rican respondents were far less individualistic than American respondents, whose scores were one of the most individualistic countries surveyed in the world. However, respondents in the Philippines also scored quite similarly to Costa Rica on this dimension, while their help-seeking behaviors were quite distinct, so values related to Individualism alone could not completely explain our results. Instead, a dimension that our observations suggested as relevant is that of Power Distance, or "the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally." The way this dimension is expressed in the classroom often relates to the relative power of the teacher to dictate student behaviors and goals among other classroom features. Hofstede's Power Distance scores in the Philippines were much higher than those of Costa Rica or the United States, and our observations indicated that students may have remained silently working at the direction of the teachers, to whom great respect is traditionally accorded. It may be that, due to interactions between underlying values

relating to Power Distance and Individualism, students from our Costa Rican site were more likely to choose to informally collaborate despite being originally told that this was individual work, while students from the Philippines site were more likely to heed the directions of the teacher and work silently and alone.

While Hofstede's framework can provide various post-hoc interpretations of the results, we should emphasize that there are many other factors that could be contributing to the different behaviors between sites, and we did not test the relationship between Hofstede's dimensions and student behaviors directly. Empirical research across a larger number of sites would be necessary to determine whether these dimensions (or others, or the use of a different cultural framework entirely) would best explain the results.

A final potential interpretation of the results is that these differences stem from some alternative source of variation such as socio-economic status or religion; or that they represent a specific culture e.g. bounded within the particular classroom, school, or city. A feature of our working definition of culture is that it is not limited by national or geographic boundaries, but rather may comprise groups with shared behaviors and values at a variety of levels of affiliation. Although in this study we choose to compare sites within three countries in part to maximize the likelihood that distinct behaviors would be present, we believe that the findings are equally relevant if the underlying factor is not national membership. Our intent was not to determine features of effective help seeking for all Costa Rican students (if such a thing were even possible), but rather to elucidate the variety of help-seeking behaviors that may be effective for learning, and address the possibility that modeling and assessing such behaviors may not be universal. We expect that future work will help to uncover generalizable causes for such variation through the use of larger samples and additional methods such as questionnaires and quantitative observational protocols.

CONCLUSION

In this study, we investigated whether a model of effective help seeking can generalize across cultures, within the context of adaptive educational software. To answer this question, we created a set of models of help seeking that try to predict learning from seventeen commonly-reported help-seeking behaviors within educational software.

The model creation process consisted of several steps. First, log data was obtained from studies of students using the same tutor on scatterplots in three countries: Costa Rica, Philippines, and U.S.A. Pre- and post-test data were also collected from the students in order to measure learning. Seventeen help-seeking features (e.g. aspects of student help-seeking behavior) were then engineered from the log files. The seventeen features were drawn from features used in past research on help-seeking behavior within educational software. After optimizing the parameters used in these features, models were developed that could predict learning in each country from a combination of features, using linear regression. Four models of effective help seeking were created, one for each of the sites in the three countries: Costa Rica, Philippines, and U.S.A, and one multi-national model created from the combined data set of the three countries. The resulting models were analyzed to understand the differences in effective help seeking between sites. The results suggested that effective help seeking looks substantially different in the Costa Rican site than in the other two countries, suggesting that while help seeking may be a cross-curricular skill, it can differ considerably between cultures.

The findings in this study indicate that there is a need for more cross-cultural comparisons of students' interactions with intelligent tutors, and the relationships between different student behaviors and learning. It will be particularly important to understand how cross-curricular skills such as help seeking differ between countries, inasmuch as there is an increased interest both in transferring educational software between countries (cf. Walker et al., 2011; Ogan et al., 2012; Nicaud, Bitta, Chaachoua, Inamdar, & Maffei, 2006), and in measuring cross-curricular skills in international comparison examinations (Beller, 2011). Simply using the same normative model in different countries may induce bias in favor of the cultural values and practices of the countries where the test developers are based, entrenching an already-present bias in education research in favor of phenomena present in developed, wealthy countries (cf. Blanchard, 2012). As such, it will be essential to replicate this type of research in additional countries; we see this step as a vital prerequisite to the wide deployment of new educational technologies and pedagogies (and related assessments). Without conducting research of this

type, we may deploy educational approaches that are inappropriate and ineffective in the new contexts where they are being used.

Acknowledgements

We thank the students who participated as subjects during these experiments. We thank REDACTED for their assistance. This work was made possible in part through a grant from REDACTED.

REFERENCES

- Arroyo, I., and Woolf, B. (2005). Inferring Learning and Attitudes from a Bayesian Network of Log File Data. In *Proceedings of the Twelfth International Conference on Artificial Intelligence in Education*, pp. 33-40.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), 716-723
- Aleven, V., McLaren, B.M., Roll, O., & Koedinger, K. (2004). Toward tutoring help-seeking: Applying cognitive modelling to meta-cognitive Skills. In J.C. Lester, R.M. Vicari, & F. Paraguacu, *Intelligent Tutoring Systems, Lecture Notes in Computer Science Vol. 3220* (pp. 227-239). Berlin Heidelberg: 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(2), 277-320.
- Al-Harathi, A.S. (2008). Learner self-regulation in distance education: A cross-cultural study. *24th Annual Conference on Distance Teaching & Learning*. Available: http://www.uwex.edu/disted/conference/Resource_library/proceedings/08_13597.pdf
- Ames, R., & Lau, S. (1982). An attributional analysis of student help-seeking in academic settings. *Journal of Educational Psychology*, 74 (3), 414-423.
- Anderson, J.R., Boyle, F.C., & Reiser, B.J. (1985). Intelligent tutoring systems. *Science*, 228, 456-462.
- Baker, R.S., Corbett, A.T., & Koedinger, K.R. (2004). Detecting student misuse of intelligent tutoring systems, In J.C. Lester, et al., *Intelligent Tutoring Systems, Lecture Notes in Computer Science Vol. 3220* (pp. 531-540). Berlin Heidelberg: Springer-Verlag.
- Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Evenson, E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., & Beck, J. (2006) Adapting to when students game an intelligent tutoring system. In M. Ikeda, K. Ashley, & T.W. Chan, *Intelligent Tutoring Systems, Lecture Notes in Computer Science Vol. 4053/2006* (pp. 392-401), Berlin Heidelberg: Springer-Verlag.
- Baker, R.S., Corbett, A.T., Koedinger, K.R., & Wagner, A.Z. (2004). Off-task behavior in the cognitive tutor classroom: when students "game the system". In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '04)*. ACM, New York, NY, U.S.A, 383-390.
- Baker, R.S.J.d., Gowda, S.M., & Corbett, A.T. (2011). Automatically detecting a student's preparation for future learning: Help use is key. In *Proceedings of the 4th International Conference on Educational Data Mining*, 179-188.

- Beller, M. (2011). Technologies in large-scale assessments: New directions, challenges, and opportunities. *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*, 25-45.
- Bielaczyc, K., Pirolli, P.L., & Brown, A.L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on Problem solving. *Cognition and Instruction*, 13(2), 221-252.
- Blanchard, E. G. (2012). On the WEIRD nature of ITS/AIED conferences: A 10 year longitudinal study analyzing potential cultural biases, In *Proceedings of the 11th International Conference on Intelligent Tutoring Systems (ITS2012)*, Chania, Greece, 280-285.
- Chi, M.T.H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanation: How students study and use examples in learning to solve problems, *Cognitive Science*, 13 (2), 145-182
- Corbett, A.T., & Anderson, J.R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 253-278.
- Csapó, B. (1999). Improving thinking through the content of teaching. In J. H. M. Hamers, J. E. H. Van Luit, & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 11–36). Lisse: Swets & Zeitlinger
- Efron, B. & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1), 36-48.
- Greiff, S. (2012). Assessment and theory in complex problem solving- A continuing contradiction? *Journal of Educational and Developmental Psychology*, 2, 49-56.
- Hofer, M., Niegemann, H. M., Eckert, A., & Rinn, U. (2006). Pädagogische Hilfen für interaktive selbstgesteuerte Lernprozesse und Konstruktion eines neuen Verfahrens zur Wissensdiagnose. [Instructional help for interactive self-directed learning processes and construction of a new procedure for knowledge diagnosis]. *Zeitschrift für Berufs-und Wirtschaftspädagogik: ZBW*, 53-67.
- Hofstede, G.H., Hofstede, G.J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind*, Third Edition. McGraw Hill.
- Karabenick, S., & Knapp, J. (1991). Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of Educational Psychology*, 83(2), 221-230.
- Kim, K., & Bonk, C.J. (2002). Cross-cultural comparisons of online collaboration. *Journal of Computer-Mediated Communication*, 8, 0.
- Koedinger, K.R., & Corbett, A.T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. *The Cambridge handbook of the learning sciences*, 61-78.
- Martin, R. (2008) New possibilities and challenges for assessment through the use of technology; In F. Scheuermann, A.G. Pereira (Eds) *Towards a Research Agenda on Computer-Based Assessment*, Ch.1.Luxembourg: Office for Official Publications of the European Communities.
- Mathews, M., Mitrovic, T., & Thomson, D. (2008). Analysing high-level help-seeking behavior in ITSs. In *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based Systems*, 312-315.
- Microsoft Corporation. (1997) *Microsoft Office 97*. Seattle, WA: Microsoft Corporation.

- Nadler, A., & Fisher, J.D. (1986). The role of threat to self-esteem and perceived control in recipient reaction to help: Theory development and empirical validation. *Advances in Experimental Social Psychology*, 19, 81-120.
- Nafukho, F.M. (2007). The place of e-learning in Africa's institutions of higher learning. *Higher Education Policy*, 20, 19-43.
- Nelson-Le Gall, S.N. (1985). Help-seeking behaviour in learning. *Review of Research in Education*, 12, 55-90.
- Newman, R.S (2002). How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. *Theory into Practice*, 41(2), 132-138.
- Nicaud, J. F., Bitta, M., Chaachoua, H., Inamdar, P., & Maffei, L. (2006). Experiments with Aplusix in four countries. *International Journal for Technology in Mathematics Education*, 13(2), 79-88.
- Ogan, A., Walker, E., Baker, R., Rebolledo, G., & Jimenez-Castro, M. (2012). Collaboration in cognitive tutor use in latin America: field study and design recommendations. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, U.S.A, 1381-1390.
- O'Lawrence, H. (2005). 'A review of distance learning influences on adult learnings: Advantages and disadvantages. *Proceedings of the 2005 Information Science and IT Education Joint Conference*, 125-135.
- Polson, M.C., & Richardson, J.J., Ed. (1988). *Foundations of Intelligent Tutoring Systems*. Lawrence Erlbaum Associates.
- Puustinen, M. (1998). Help-seeking behavior in a problem-solving situation: Development of self-regulation. *European Journal of Psychology of Education*, 13(2), 271-282.
- Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological methodology*, 25, 111-164.
- Ramsey, F., & Schafer, D.W. (1993). *The statistical sleuth: an intermediate course in statistical methods*. Independence, KY: Brooks Cole Publishing.
- Rodrigo, M.M.T., Baker, R.S.J.d., & Rossi, L. (in press) A comparison of student off-task behavior in computer-based learning in the U.S.A and the Philippines. To appear in *Teachers College Record*.
- Rodrigo, M.M.T., Baker, R.S.J.d., Agapito, J., Nabos, J., Repalam, M.C., Reyes, S.S., & San Pedro, M.C.Z. (2012) The effects of an interactive software agent on student affective dynamics while using an intelligent tutoring system. *IEEE Transactions on Affective Computing*, 3 (2), 224-236.
- Ryan, A.M., Hicks, L., & Midgley, C. (1997). Social goals, academic goals, and avoiding seeking help in the classroom. *Journal of Early Adolescence*, 17(2), 152-171.
- Ryan, A.M., Pintrich, P.R., & Midgley, C (2001). Avoiding seeking help in the classroom: Who and why? *Educational Psychology Review*, 13(2), 93-114.
- San Pedro, M.O.C., Baker, R.S.J.d., & Rodrigo, M.M. (2011). Detecting carelessness through contextual estimation of slip probabilities among students using an intelligent tutor for mathematics. In G. Biswas, et al., *Artificial Intelligence in Education, Lecture Notes in Computer Science, Vol. 6738/2011* (pp. 304-311). Berlin Heidelberg: Springer-Verlag.
- Shih, B., Koedinger, K., & Scheines, R. (2008). A response time model for bottom-out hints as worked examples. In *Proceedings of the 1st International Conference on Educational Data Mining*, 117-126.
- Spencer-Oatey, H. (2008) *Culturally Speaking. Culture, Communication and Politeness Theory*. 2nd edition. London: Continuum.

- Stanton-Salazar, R.D., Chavez, L.F., & Tai, R.H. (2001). The help-seeking orientations of Latino and non-Latino urban high school students: a critical-sociological investigation. *Social Psychology of Education, 5* (1), 49-82.
- Taplin, M., Yum, J.C.K, Jegede, O., Fan, R.Y.K., & Chan, M.S.C. (2001). Help-seeking strategies used by high-achieving and low-achieving distance education students. *Journal of Distance Education, 16* (1), 56-69.
- VanLehn, K., Jones, M.R., & Chi, M.T.H. (1992). A model of the self-explanation effect. *Journal of the Learning Sciences, 2* (1), 1-59.
- Vatrapu, R.K. (2008). Cultural considerations in computer supported collaborative learning. *Research and Practice in Technology Enhanced Learning, 3*, 159-201.
- Walker, E., Ogan, A., Baker, R.S.J.d., de Carvalho, A., Laurentino, T., Rebolledo-Mendez, G., & Castro, M.J. (2011). Observations of collaboration in cognitive tutor use in Latin America. In *Proceedings of the 15th International Conference on Artificial Intelligence in Education, 575-577*.
- Wood, H., & Wood, D. (1999). Help-seeking, learning, and contingent tutoring. *Computers & Education, 33*, 153-169.
- Woolf, B.P. (2009). *Building Interactive Intelligent Tutors*. Burlington, MA: Elsevier, Inc.