The Influence of School Demographics on the Relationship Between Students’ Help-Seeking Behavior and Performance and Motivational Measures

Shamya Karumbaiah
Penn Center for Learning Analytics
3700 Walnut St
Philadelphia, PA 19104
+1 (877) 736-6473
shamya@upenn.edu

Jaclyn Ocumpaugh
Penn Center for Learning Analytics
3700 Walnut St
Philadelphia, PA 19104
+1 (877) 736-6473
jlocumpaugh@gmail.com

Ryan S Baker
Penn Center for Learning Analytics
3700 Walnut St
Philadelphia, PA 19104
+1 (877) 736-6473
ryanshaunbaker@gmail.com

ABSTRACT
Demographic information often proves useful for finding subpopulations in educational data. Unfortunately, it is often not collected in the log files of online learning systems, which serve as one of the primary sources of data for the Educational Data Mining community. Recent work has sought to address this issue by investigating school-level differences in demographics, which can be used to discover trends in data where individual-level variation may be difficult or impossible to acquire. In this study, we use this approach to investigate the effect of demographic patterns on hint usage in an elementary level mathematics system, comparing this use to performance and motivational measures. In doing so, we expand upon the research into help-seeking behaviors, which typically takes a cognitive approach. Our results suggest the need to better understand what social factors are most likely to motivate help-seeking behaviors, particularly since research on their effectiveness has been mixed.

Keywords
Help-seeking, Demographics, Math Learning, Math Identity, Self Concept, Self Regulated Learning

1. INTRODUCTION
Many studies into complex constructs like motivation, interest, and engagement involve either small-scale experiments or larger convenient samples of middle-class, undergraduate students (see discussion in [21]), which can make it difficult to determine the extent to which these findings will generalize to new populations of students. This is often due to the practical constraints of research projects involving the budget, recruitment, accessibility and time required to acquire the level of detail used in these studies. However, this trade-off sometimes means that conclusions are not replicated across broad demographic contexts. Conversely, Educational Data Mining (EDM) researchers often have larger sample sizes than are seen in experimental psychology, for example, but the source of typical EDM data (i.e., intelligent tutoring systems) often limits the practicality of obtaining demographic variables from individual students.

Shamya Karumbaiah, Jaclyn Ocumpaugh and Ryan S Baker "The Influence of School Demographics on the Relationship Between Students’ Help-Seeking Behavior and Performance and Motivational Measures" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 99 - 108

Beyond practicality (e.g., the ease of acquiring log-file data on student interactions compared to student demographic data), there are sometimes other legitimate concerns, including those related to student privacy. For example, even when a partner school or university has documented the demographics of individual students, their release to a researcher increases the risk of potentially re-identifying students, particularly in rural parts of the country where the analysis of say, the seven children of a minority ethnic group in a small school narrows the potential matches for sensitive information considerably.

Yet considerable research shows that demographic factors are often related to differences in educational outcomes more generally (see [14]) and to constructs related to motivation more specifically [50, 50, 44]. This suggests that researchers in the EDM community should make greater efforts to overcome the challenges involved in collecting demographic data in order to ensure population validity (e.g., [29]). As such, some researchers within the EDM community have sought to extend student learning models to include information from the broader context, building models at the class, school-, school-cluster level instead of just the student-level [47, 31]. [49] used school-level demographics and students’ prior performance to cluster schools into groups, improving model performance.

This study uses this approach to investigate hint-usage, incorporating the broader demographic context into the investigation while also answering a call to pay greater attention to the social factors influencing help seeking behaviors after a recent review of the literature found that their effectiveness was highly variable [15]. We do so in the context of Reasoning Mind—an Intelligent Tutoring System for elementary mathematics—where we explore how readily-available school-level demographics might reveal how hint usage correlates to measures of student performance and motivation (i.e., mathematics self-concept).

2. PRIOR WORK
Help - mostly in the form of on-demand, contextual, real-time hints - is a common feature in most Intelligent Tutoring Systems (ITSs) [46] and has long been believed to foster emerging concepts and principles in a student’s learning [7] and support struggling students during problem-solving [1]. Yet help-seeking behaviors are not always beneficial [1, 2, 4]. While much of the prior work on help-seeking in ITSs has focused strictly on its cognitive effects, other research suggests that social factors may influence these patterns.

2.1 Help-Seeking: A Cognitive Lens

The literature on help-seeking behaviors in ITS now stretches back over twenty years (see extensive review in [4]). As it quickly became apparent that the availability of hints did not ensure their effective use, work began to identify the factors that led to a positive relationship between help-seeking behaviors and student learning.

In one of the earliest studies, Anderson et al. [6], compared the use of explanatory hints and so-called bottom-out hints (which simply provided the student with the correct answer and found that neither hint type was correlated with learning. In part, this may have been due to selection bias. That is, hint usage is typically a sign of struggling students, who often do not make substantial learning gains (see discussion in [4]).

After early discoveries of a negative correlation between hint usage on student learning in one context [1], researchers began to develop a taxonomy of maladaptive help-seeking behaviors—including categories like help abuse (the overuse of help) and help avoidance (the underuse of help)—was also developed [13]. Most studies analyzed the effectiveness of hints by focusing on the relationship between help-seeking behavior(s) and student outcome(s), with some researchers emphasizing that the intentionality of help-seeking behavior makes it a good candidate for understanding students' self-regulated learning (SRL) strategies [4, 16].

A number of studies have attempted to identify the degree of help needed at any given moment (e.g., Koedinger & Aleven's [24] *assistance dilemma*), and experimental results have resulted in notable findings. For example, (1) on-demand hints led to greater learning gains than automatic hints in middle-school mathematics [32]; (2) hint content (goal feedback versus other kinds of feedback) is related to student learning in Geometry [27]; (3) hints about which step to try next to improve student learning in logic proof [36].

In general, the review of the literature suggests that increasing hint usage does not always lead to better domain-level learning [4]. However, the EDM literature on help-seeking in ITSs has produced research which aggregates into a complicated and contradictory narrative, including: (1) a negative association between hint usage and learning [2]; (2) a positive association between hint usage and learning [11, 48]; (3) a positive association between hint usage and learning only when time per hint level is considered [25] or when adaptive versus maladaptive help-seeking is differentiated [3]; (4) a positive association between time spent in bottom-out hints and learning [35]; (5) a negative association between the number of bottom-out hints used and learning [26]; (6) positive benefits for students but only when they have a medium level of skill [33]; (7) a negative association between help avoidance and learning early within practice [5] and on a transfer post-test [9].

In addition, individual differences in self-regulation were observed in how students process hints and how that impacts their performance [16]. Vaessen et al [45] found that students' achievement goals (mastery and performance goals) are closely related with their help-seeking and could be used to predict their strategies for help-seeking. Despite a considerable volume of research, the effectiveness of help-seeking remains an open question—and the clearest thing that we can say is that the relationship between hint usage and learning is complicated.

2.2 Help-Seeking: A Social Lens

While the role of social factors on help-seeking behaviors has not been the primary focus of the EDM community (see [4]), the social evaluation of help-seeking behaviors is well established in the literature. For instance, some learners may feel that asking for help is either a sign of incompetence or a challenge to their autonomy [39]. Likewise, Howley et al. [18] suggests that asking for help may trigger experiences of evaluation anxiety – the fear of being judged.

These kinds of concerns seem ripe for socio-cultural variation, and a few studies have begun to explore how these differences may emerge. For example, Tai et al. [38] increased students' help-seeking behaviors by changing the way they labeled those actions within the system. That is, they started by referring to the ITS as the students' teammate, and the designed the system so that students who needed help could choose to "work together" with the system. This adaption apparently reduced the ego-threat related to admitting a lack of knowledge (e.g., [39]) and improved student learning.

Other studies have specifically investigated demographic differences in help-seeking behaviors. Ogan et al. [30] found that the EDM models on effective help-seeking did not transfer well between countries (namely Costa Rica, the Philippines, and the USA). Likewise, Arroyo et al. [8] found that the effectiveness of different hint designs varied by gender. Specifically, girls benefited more from highly interactive hints, while boys did better with less interactive hints.

Thus, there is a need for more research to look at social factors while studying help-seeking. Such studies should focus on students' broader context to understand under what circumstances lead to desired student outcome. In this paper, we study students' help-seeking behavior in an online math tutor used in traditional classrooms during regular instruction. Given the context in which the students use this ITS, we focus on school as the social context and analyze the influence of school demographics on the relationship between student outcomes (math performance and math self-concept) and their help-seeking behavior. We aim to shift the focus of help-seeking research in the EDM community from purely cognitive factors to the contextual factors that might play a more prominent role than is assumed.

2.3 The Role of Demographics in Predicting Student Outcomes

This section summarizes prior work on the role of demographics in the student outcomes of interest in this study - math performance and math self concept.

2.3.1 Demographics and Math Performance

The literature addressing demographic differences in learning outcomes (at least in a U.S. context) is now so vast that it would be difficult to review even if it were limited to a single domain (e.g., mathematics). Once referred to as the achievement gap, more and more scholars are now discussing it in terms of an opportunity gap, as findings generally show that achievement patterns favor groups for whom the educational system was initially designed.

Scholars point out that reframing this discussion in terms of opportunities to learn emphasizes the need to address the environmental inadequacies that may occur. Childs [14] analysis shows, for example, that minority students are just as likely to value mathematics, but are less likely to attend schools where advanced mathematics classes are offered.

However, less tangible differences may also play a role. For example, if students' patterns of communication are different than those expected by educators (e.g., [19]), their attempts at help-seeking may not receive adequate uptake. Such experiences could discourage students from future help-seeking behaviors, although one could imagine that the ability to get help from an ITS could also mitigate this reluctance.

2.3.2 Demographics and Self-Concept

Demographic variables have also been shown to correlate with motivational constructs, like math self-concept. Math self-concept (sometimes used interchangeably with self-efficacy, although see [12]) has been found to be a predictor of various measures of achievement and career choice (see [13]). It has also been linked to motivational constructs, including achievement goal orientation, anxiety, and self-concept [34].

Early work proposed that self-efficacy was a product of a person's own accomplishments and the feedback they receive on their work [10, 43]; however, more recent studies have indicated that the source of self-efficacy may vary along demographic lines like gender and ethnicity [50, 50, 44]. For example, Klassen [23] investigation of self-efficacy among seventh grade students found that ethnic majority students followed Bandura's predictions, citing personal achievements as a source of self-efficacy, but ethnic minority students were more likely to cite social factors. Else-Quest, et al. [15] studied the intersection of gender, ethnicity, and achievement in 10th grade students from a large northeastern city and found that males reported greater math self-concept and expectation of success as compared females, but no gender differences across ethnic groups were found.

Other research on self-efficacy suggests that it is malleable and can be influenced by social agents [51], and there are significant efforts to understand how to support underrepresented groups, who may struggle against implicit stereotypes on top of normal learning struggles as their domain knowledge matures [37]. Previous research shows that scores on social identity ratings (e.g. gender and cultural identity ratings) peak when people are experiencing uncertainty [17]. This could suggest that students could become more susceptible to negative cultural stereotypes (e.g., [37]), particularly those related to STEM performance, during periods of confusion associated with learning, making help-seeking an important behavior to study for its associations with self-concept.

Given these findings, it seems likely that self-concept could vary not just by the demographics of individual students, but also based on how those demographics influence the cultural interactions at a school level. That is, in a school where larger numbers of students share a particular demographic characteristic, we might see help-seeking behaviors that emerge as a reflection of the practices more typical of that group.

3. DATA COLLECTION

3.1 Reasoning Mind

This study analyzes data from students using Reasoning Mind (RM) Foundations (Figure 1), an intelligent tutoring system for elementary mathematics, produced by Imagine Learning. It currently serves over 100,000 U.S. students annually. The majority of these students are in Texas, but they represent a range of traditionally underrepresented populations across rural, urban, and suburban schools. Key components of this system include socio-technical innovations, including those that are designed to directly support teachers [21] and those that are designed to mimic other social experiences in the classroom, including both virtual peers and the signature pedagogical agent, known as the Genie, that guides students in their learning.

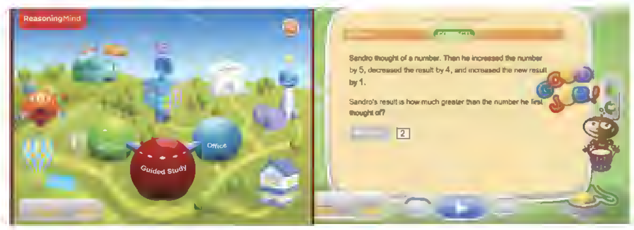

Figure 1. Reasoning Mind Foundations home screen (*left*) and an example problem (r*ight*)

In this blended environment, students learn through self-paced problem solving, interactive explanations, and skill-based games. Problem sets are classified into three groups based on increasing levels of difficulty: (1) A-level problems on fundamental skills; (2) B-level (optional) problems on a combination of skills; and (3) C-level (optional) problems on higher-order thinking skills. Our past study [20] suggests a close relationship between inconsistencies in students' math performance and their math self-concept – the two student outcomes studied in this paper. Reasoning Mind Foundations is generally used in traditional classrooms. Teachers assign/unlock problem sets for students based on the topic of instruction. Past studies of Reasoning Mind Foundations have shown high student and teacher acceptance, increases in test scores, high time on task, and a positive affective profile [21].

3.2 Hints in Reasoning Mind

Hints are an integral part of Reasoning Mind Foundations. These are delivered only on student request and contains conceptual feedback intended to help students solve the problem. Figure 2 demonstrates a hint in the system for one of the basic A-level problem in Reasoning Mind Foundations. They are multi-level and do not always contain a bottom-out hint.

Figure 2. *Top* - Problem screen with a button to view hint (highlighted in green) *Bottom* - Hint displayed to the student when they request to view

3.3 Participant Schools

In order to ensure consistency in the type of data used to characterize schools, this study limits itself to Reasoning Mind schools that fall under the purview of the Texas Education Agency's (TEA) and further filters out schools where less than 25 students were using the software to avoid noise in the correlations reported below. This resulted in data from 110 Texas schools across 25 school districts who used Reasoning Mind during the academic year 2017-2018 as part of their regular mathematics instruction. There are a total of 9,122 2^{nd} through 5^{th} grade students in this data (4,749 2^{nd} graders, 1,964 3^{rd} graders, 1,582 4^{th} graders, and 827 5^{th} graders). On average, there were 75 students using Reasoning Mind Foundations per school (min = 25; SD = 70) and 364 per school district (SD = 730), with one large urban district in Texas constituting the majority of our data, with 3,039 students across 62 schools.

Comprehensive log data captured student interactions with the system for the entire period, resulting in data for all 9,122 students. Surveys were administered once at the beginning and once at the end of the year to collect data on student math identity, resulting in complete surveys for 2,238 students.

4. DATA EXPLORATION

Considerable variation exists in the measures being analyzed in this study: help-seeking behaviors (i.e., hint usage), math performance, and pre- and post-year measures of math self concept.

4.1 Exploring Help-Seeking

From the interaction log data, we operationalize help-seeking behavior using as the number of hints used by a student in Reasoning Mind Foundations. As shown in Figure 3 (left), students in this study averaged less than 30 hint requests annually (mean = 27.01, SD = 55.72).

4.2 Exploring Math Performance

For the purposes of this paper, math performance is defined as the accuracy of student responses to A-level problems in Reasoning Mind Foundations. Accuracy on these problems, which represent the core curriculum within the software, is computed from the interaction log data. As presented in Figure 3 (right), student-level calculations show a mean of 0.77 (SD = 0.14).

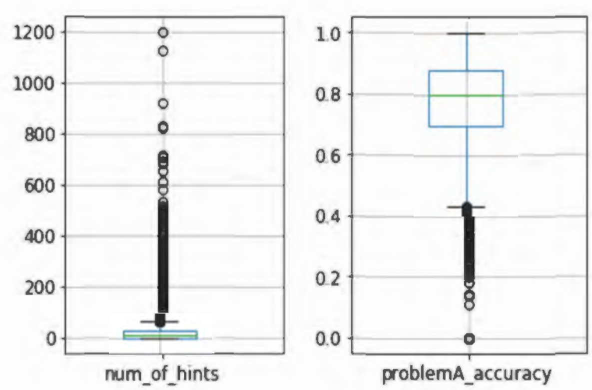

Figure 3. Distribution of the number of hints (*left*) and math performance (accuracy in A-level problems; *right*). The green line in the box indicates the median value.

4.3 Exploring Math Self Concept

Students' self concept in mathematics was measured using a five-item survey adapted from Marsh et al. [28]. This survey was administered twice--once at the beginning of the academic year (pre) and once at the end of the academic year (post). The survey included questions like *Math just isn't my thing; Some topics in math are just so hard that I know from the start I'll never understand them.* Students took the survey voluntarily, and each item in the survey was answered with a four-point Likert scale.

This study analyzes survey responses from 2,238 students across 22 Texas schools. The distribution of students' responses is given in Figure 4 (self concept pre: mean = 2.64 standard deviation = 0.77; self concept post: mean = 2.44, standard deviation = 0.80). The internal consistency of these items was found to be satisfactory with a Cronbach's α of 0.74.

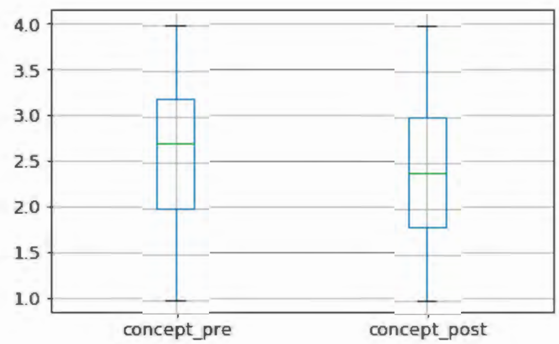

Figure 4. Distribution of number of pre and post measures of math self concept.

4.4 Exploring School-Level Differences

Next, we explored the school-level differences in student outcomes (math performance and self concept) and hint usage. As we can see in Figure 5 and Table 1, there is considerable variance in the variable aggregates (mean) across the schools, especially in hint usage and math performance.

Table 1. Mean and standard deviation (SD) of the school-level aggregates of the variable and outcomes

	Mean	SD
Hint Usage	24.5	21.3
Math Performance	0.78	0.04
Math Self Concept (Pre)	2.69	0.35
Math Self Concept (Post)	2.43	0.15

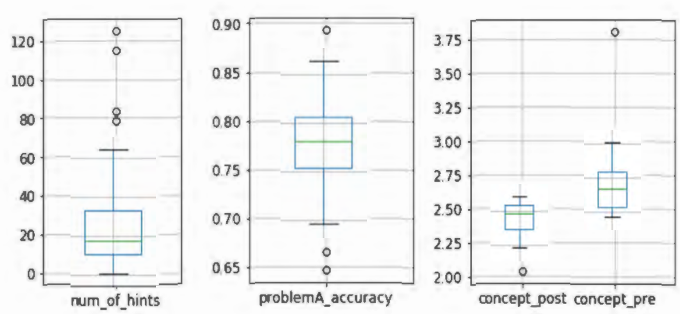

Figure 5. Distribution of school-level aggregates of the variable and outcomes.

4.5 Summarizing School-Level Demographics

We characterize the schools in our sample using demographics from the Texas Education Agency's (TEA) public data repository. These data capture the contextual factors that are likely to affect the school culture or climate and defines the social context in which students using RM Foundations.

Table 2. Mean and standard deviation (SD) of the school-level demographics. EcD - Economically Disadvantaged; LEP - Limited English Proficiency; SpEd - Special Education

	Mean	SD
Urbanicity (binary)	0.6	0.49
% EcD	78.3	16.6
% LEP	41.4	20.6
% SpEd	6.9	3.1

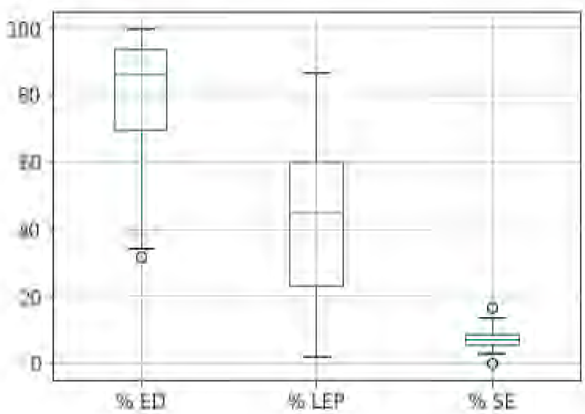

Figure 6. Distribution of percentages of school-level demographics for the 110 schools selected in this study. ED - Economically Disadvantaged; LEP - Limited English Proficiency; SE - Special Education

Table 2 summarizes the first set of school-level demographics obtained from TEA sources, including the percentage of students at the school who are classified as (1) Economically Disadvantaged (EcD), as (2) Special Ed (SpEd) or as (3) Limited English Proficiency (LEP), as well as (4) the urbanicity of the school. These terms are defined by the State of Texas as follows [40]. Students are classified as EcD if they qualify for free or reduced-price meals under the National School Lunch and Child Nutrition Program, and it is worth noting that a large proportion (avg = 40%) of Texas public school students qualify for this status [40]. SpEd classifications are given to students who qualify for services for cognitive, emotional, or physical disabilities. LEP status is conferred for students whose primary home language is not English and who also fail to meet proficiency standards as established by either an approved testing measure or by a Language Proficiency Assessment Committee (LPAC). Finally, the TEA classifies a school district as urban (or not) [41] based on whether its school district (a) is located in a county with a population of at least 960,000; (b) has the largest enrollment in the county or its enrollment is greater or equal to 70% of county's largest district. As seen in Table 2 and Figure 6, we have a diverse set of schools along these dimensions.

We also considered school-level data on the percentage of students representing major ethnic/racial groups. As Table 3 shows, Hispanic students are by far the largest group in these schools (mean = 63.5%), followed by African American students (mean = 17.5%), White students (mean = 13.5%) and then Asian students (4.5%), but as Figure 7 the schools show considerable variance in terms of this composition. To avoid noisy results, this analysis considers only groups that constitute at least 5% of the student population: Hispanic, African American, White and Asian.

Table 3. Standard deviation (SD) of the school-level percentages of ethnicities. Categories constituting less than 5% of the data were excluded from further analysis.

	Mean	SD
% Hispanic	63.5	24.5
% African American	17.5	17.8
% White	13.1	16.2
% Asian	4.5	7.8
% American Indian*	0.36	0.4
% Pacific Islander*	0.04	0.1
% Two or More Races*	1	1

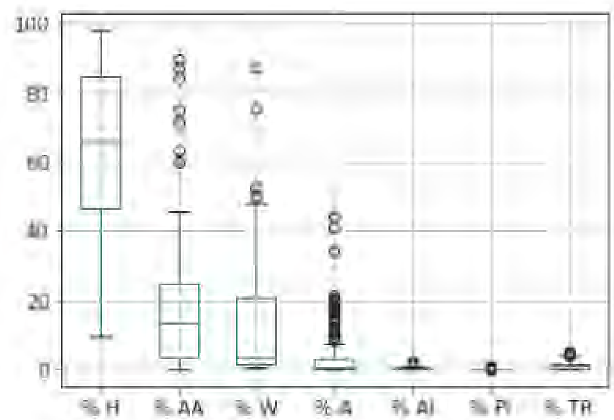

Figure 7. Distribution of percentages of school-level ethnicities for the 110 schools selected in this study. H - Hispanic; AA - African American; W - White; A - Asian; AI - American Indian; PI - Pacific Islander; TR - Two or races

5. ANALYSIS

Our data exploration (Section 4) suggests that help-seeking behavior, math performance, math self-concept, and demographics each vary by school. Thus, we conduct a two-step data analysis to explore how help-seeking behavior might differ based on student demographics, while controlling for performance and motivational measures.

In the first step, we determine how closely students' math performance (and self concept measures) correlate to their hint usage, within each school, using Spearman ρ correlations due to non-normality in the data. That is, we produce three new measures for each student, the correlation between hint use and performance on A-level problems, the correlation between hint use and the pre-year survey of self concept, and the correlation between hint usage and the post-year survey of self concept.

In the next step, we determine whether the differences in these correlations are themselves correlated to school-level demographics. Note that in the first step, the unit of analysis for the correlations is the student, but in the second step, the unit of analysis is the school. We conduct two-tailed tests to report the significance levels.

6. RESULTS

6.1 Help-Seeking and Student Outcomes

Figure 8 summarizes the results for our study, showing the distribution of correlations across schools between students' hint usage and their math performance and math self-concept (taken once at the beginning (pre) and again at the end of the year (post)).

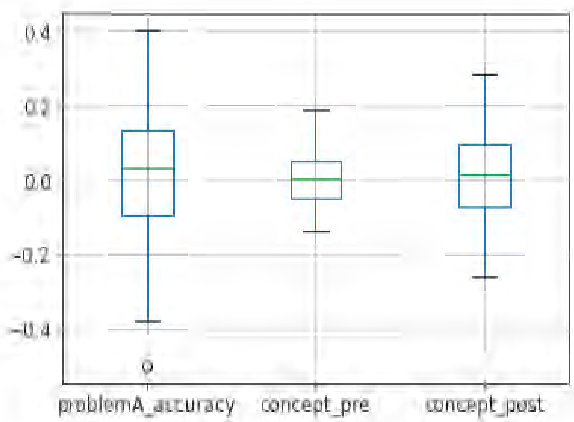

Figure 8. Distribution of correlations across schools between students' hint usage and outcomes.

6.1.1 *Help Seeking and Math Performance*

Clustering students by schools allows us to see that the relationship between hint usage and math performance differs in ways that might be missed if this aggregation were not used. This is true even when demographic descriptions are not used to describe the data.

Specifically, when student measures are aggregated at the school level, we see that the correlation between hint usage and math performance ranges from -0.39 to 0.40 (SD = .18). In contrast, when we do not aggregate students into school-level populations (instead treating them all as a single population), there is not a significant relationship between hint usage and math performance ($\rho = -0.008$, $p = 0.44$).

6.1.2 *Help Seeking and Math Self Concept*

Like math performance, math self concept also shows signs of sub-population differences. When students are aggregated into school-level populations, the correlations between hint usage and math self-concept show a relatively wide range.

For pre-year surveys, the correlation ranges from -0.14 (students with lower self-concept are most likely to use hints) to 0.19 (students with higher self concept are most likely to use hints), and an even wide range is found for post-year survey correlations (-0.27 to 0.30). In contrast, when the students in this data were treated as a single population, the correlations were non-significant ($\rho = -0.008$, $p = 0.442$) for pre and $\rho = -0.007$, $p = 0.77$ for post).

6.1.3 *Summary of Help Seeking Variance*

There is considerable variance in the school-level correlations between hint usage and student outcome measures (SD = 0.18 for math performance, SD = .084 for pre-year self concept, SD = 0.118 for post-year self concept). This variance indicates that students likely have different motivations for using hints, and they be more effectively used by some student populations than by others.

As seen in Figure 8, the median of the correlations is centered close to zero. For these schools, there is no association between hint usage on student outcomes. Figure 8 also shows that the distribution of these correlations is not skewed, meaning that hint usage is not universally positively or negatively associated with student outcomes across schools.

The schools at the tail ends of these distributions are interesting case studies. They represent the cases where hint usage has either a notably high positive correlation or a notably high negative correlation with our outcome measures. In schools where there is a high positive correlation between these variables, the use of hints appears to be beneficial, but the converse is true for those schools that have high negative correlations. As such, it becomes important to understand what demographics are involved in order to address any potential disparate impacts of the hint function in the system.

6.2 The Influence of School Demographics

School-level demographic variables help to capture some of the variance in the relationship between hint usage and the student outcomes measured in this study (math performance and math-self concept). These findings are summarized in Tables 4 and 5.

Table 4. Correlations between school-level demographics and the correlations resulted between students' math performance and interaction features. p-value in parenthesis. Significant correlations in bold.

	Correlation between number of hints and		
	Math performance	*self concept Pre*	*self concept Post*
Urbanicity	**0.292** (0.002)	0.130 (0.564)	0.080 (0.729)
%EcD	**0.256** (0.007)	0.182 (0.417)	-0.288 (0.205)
%LEP	**0.314** (0.001)	**-0.452** (0.035)	**-0.565** (0.008)
%SE	-0.002 (0.982)	**0.463** (0.030)	**0.444** (0.044)

Table 5. Correlations between school-level ethnicity and the correlations resulted between students' math performance

and interaction features. p-value in parenthesis. Significant correlations in bold.

	Correlation between number of hints and		
	Math performance	*Self concept Pre*	*Self concept Post*
% Hispanic	0.094 (0.329)	0.123 (0.587)	-0.153 (0.507)
% African American	0.054 (0.579)	-0.260 (0.243)	-0.174 (0.451)
% White	-0.194 (0.042)	0.103 (0.647)	0.095 (0.683)
% Asian	-0.037 (0.703)	-0.071 (0.753)	-0.107 (0.644)

6.2.1 School-level Demographics, Help Seeking, and Math Performance

As Table 4 (above) shows, the relationships between hint usage and math performance differ significantly in terms of the school's urbanicity (ρ = .292, p = .002) as well as differences in the percentage of students who are economically disadvantaged (EcD; ρ = .256, p = .007) and limited English proficiency (LEP; ρ = .314, p = .001). Specifically, the association between higher hint usage and math performance is positive among students from urban schools than students from rural or suburban schools. Schools with higher percentage of students who are economically disadvantaged (EcD) or limited English proficiency (LEP), show the same trends. Conversely, rural and suburban schools show an inverse relationship between hint usage and math performance, suggesting that these students may be using hints ineffectively.

However, as Table 5 shows, other demographic categories that are often considered in educational research, namely ethnicity, are not particularly useful in this context. Schools with smaller populations of White students are more likely to show a positive relationship between hint use and math performance, whereas this relationship is more likely to be negative in schools with larger populations of White students. However, neither the percentage of African American students (which tends to be relatively small in the state of Texas and in this sample in particular) nor the percentage of Hispanic students (which tends to be quite large) is correlated with this relationship.

6.2.2 School-level Demographics, Help Seeking, and Math Self Concept

School-level demographics are less helpful in explaining the relationships between hint usage and math self concept. The relationships between hint usage and math self concept differ significantly in terms of the percentage of students with limited English proficiency (ρ = -.452, p = .035 for pre; ρ = -.565, p = .008 for post), and the percentage of students in special education (ρ = .463, p = .030 for pre; ρ = .444, p = .044 for post)). Specifically, in schools that serve a higher percentage of LEP students, there is a negative correlation with hint usage and self concept. Whereas, hint usage is more common among students with high self concept in schools that serve fewer LEP students. This finding is somewhat stronger for the end of year surveys than the start of year surveys. The opposite pattern is shown among schools that serve a higher percentage of SpEd students. In these schools, there is a positive correlation between hint usage and self concept, where as that relationship is negative in schools that serve fewer SpEd students. This relationship is consistent across the start of the year and end of the year surveys.

Other demographic factors from Table 4 that were predictive of the relationship between help seeking and math performance, namely urbanicity and EcD, were not significant for the relationship between help seeking and math self concept. School-level descriptions of ethnicity (Table 5) also did not help to explain the variance between math self concept and hint usage.

7. DISCUSSION

7.1 Overview of Results

Hint-seeking behaviors have been a source of interest among EDM researchers since the early days of the field, yet understanding which hints are effective, to whom, and under what conditions remains a somewhat elusive task.

A large part of answering these questions likely lies in understanding what motivates a student to seek help. Ideally, we would like students to use these functions to improve their understanding of the material, but as these results show, students who are struggling do not always make use of available resources (e.g., in schools where low performers are not requesting as many hints).

However, within this data—which studies students in the same state using the same mathematics learning system—there are also schools where low-performing students are requesting lots of hints. If these students are benefiting from this hint usage, it is not measurable with the variables considered in this study. This finding suggests that the hints could be less effective at helping these particular students to learn the material.

At least part of this variance seems to be related to school-level demographics, but interestingly, the schools where hint usage appears to be most advantageous are those that enroll larger numbers of students who would typically be thought of as disadvantaged by the school system. That is, schools with fewer LEP students are more likely to have low performers who are requesting lots of hints. Schools with fewer students receiving free or reduced price lunch are more likely to have low performers who are requesting lots of hints. Schools in large urban centers are less likely to have low performing students who are requesting lots of hints.

The relationship between hint usage and self-concept is also complicated. Students in schools that serve more LEP students tend to show a negative relationship between self concept and hint usage. That is, those students who are unsure of themselves are asking for more hints (in those schools). However, in schools that serve more SpED students, the relationship between self concept and hint usage is negative. It is also possible that the smaller number of students sampled for self concept (compared to math performance) made it more difficult for these relationships to emerge.

Ethnic population differences were not particularly revealing in this study, and it is not entirely clear why. It is possible that, say, the LEP findings are strong enough to warrant further divisions to the subpopulations included in this study, a possibility that has not yet been explored in this data. However, it is also possible that some of the linguistic differences that influence classroom practices different ethnic groups within the United States (e.g.,

[19])—practices that may include figuring out how to ask for help—are less relevant in an online context like Reasoning Mind where the student is simply pressing button to request a hint.

Gender was not investigated in the current paper, as public schools generally have balanced gender distributions (as was the case in this dataset), leading to limited power to observe any difference that might exist. This leads to a more general point. It would be beneficial to analyze the impact of demographics at the student level, both to replicate the relationships seen here and to study whether students who are outliers in their own schools have different patterns. However, collecting student-level data is not always feasible, and this study has demonstrated that school-level aggregates can likely improve our efforts to better capture variance in help-seeking behaviors.

7.2 Implications to ITS Designers

One of the main implications of this paper to ITS designers is that a universal design that focuses on improving student outcomes while ignoring individual or group differences might not be a fair design consideration. Personalization in help-design has primarily focused on student cognition to provide "in-context" hints based on the pedagogical content. We suggest expanding the definition of "in-context" to include broader contextual factors that impact student outcomes.

To illustrate this, let's take the example of LEP. As stated in Section 6.2, there is an inverse influence of help-seeking and the two student outcomes (performance vs. self concept). In schools with a higher percentage of limited English proficient students, higher hint usage is associated with high math performance but low math self-concept. On the other hand, in schools with more native English speakers, higher hint usage is associated with low math performance but high math self-concept. This is an interesting case of conflict for ITS designers to investigate further. Is the text-heavy nature of the hints contributing to this finding? Is it that while limited English proficient students use hints to improve on their math skills, the cognitive load in processing more verbal content is causing a negative impact on their self-efficacy? Such investigations could open up opportunities for design innovations to better support students. Would it help to use multiple representations (visual, auditory, symbols) and give autonomy to the students to make the choice? In summary, including school-level demographics to the analysis of complex constructs like help-seeking is an important step in appropriately situating the design decisions to the student context.

7.3 Limitations and Future Work

We acknowledge that there are other socio-cultural aspects that influence a student's engagement and learning with an ITS. In the case of students' help-seeking behavior, the perceptions of help-seeking within their classroom (peers, teachers) and outside (family, friends) can be influence student choices. While this paper focuses on broadly-defined school-level demographics, we believe that it would be beneficial to look at other influencers from the student's social context. For instance, the pedagogical practices of the teacher in the math classroom could influence what students perceive as appropriate help-seeking.

Within Reasoning Mind Foundations, specifically, teachers are able to explicitly choose to design which problem types to assign to their students, in line with their pedagogical goals and perceptions of the appropriate level of difficulty. These choices likely reflect the classroom culture they are hoping to foster—including the degree to which they encourage students to attempt new skills and persevere in the face of challenges. There is an opportunity to explore this data to study the impact of teacher choices on the relationship between help-seeking and student outcomes.

More broadly, the priorities of the school district and state might also impact the pedagogical choices made in schools. Teachers' choices are influenced by public policy. Shortly after the completion of our data collection, Texas issued letter grades (A-F) [42] to its school districts based on a complex formula involving overall student performance on standardized exams, overall year-to-year improvement, and improvement for specific sub-groups. These ratings were generally lower in districts with higher rates of economically disadvantaged students, creating different degrees of pressure where demographics differ. The pressure of performing well (as measured by standardized tests), in many cases with limited resources, could influence what is being prioritized as the goal of math learning in these schools. While quantifying these factors to include in an analysis is not straightforward, these factors no doubt drive the type of differences that are seen between schools with different demographics.

8. CONCLUSION

Self-regulation is an important aspect of successful learning. Intelligent Tutoring Systems like Reasoning Mind Foundations provide a unique opportunity for students to practice self-regulation by taking control over their choices in the learning environment. Help-seeking is a particularly relevant SRL process within this type of learning system, given the prominence of hints in ITSs. In this paper, we demonstrate that school-level demographics can have a significant influence on the relationships between students' help-seeking behavior and student outcomes. In doing so, we question the prevailing assumption that complex constructs like help-seeking can be considered without also considering student context. This calls for greater consideration within our field of social, cultural, and economic influencers outside cognition.

Amidst the mixed results from empirical studies on the effectiveness of hints, Aleven and colleagues [4] continue to recommend the use hints in ITSs and suggest making four key methodological distinctions when studying interventions designed to promote help-seeking - (1) effects on learning in the same learning environment versus a new environment; (2) effects on current learning versus future learning; (4) effects on learning in the same domain versus another; (3) effects on SRL process versus domain-level learning. We propose to extend upon the list of these methodological considerations, suggesting that researchers also (5) explore the effects of help-seeking designs in one demographic context versus another. We make this proposal while fully understanding both the practical challenges elaborated in the introduction and the definitional issues elaborated in Section 7. However, as we can see that such demographic effects are present even within a single U.S. state (albeit one of the larger and more diverse U.S. States), it is worth considering the ways in which different groups of people may attach different meanings to the behavior of help seeking. In particular, research should consider the ways in which help-seeking might be interpreted as an imposition or as an admission of failure, since, as we discussed in Section 2, these interpretations likely vary from one culture to another. By considering demographics in our research on help-seeking—and on SRL in general—we increase the likelihood that our findings will apply to the full diversity of learners using ITSs and related systems today.

9. ACKNOWLEDGMENTS

Our thanks to the NSF (Cyberlearning Award #1623730) for sponsoring this project, and our thanks to Matthew Labrum and Wanjing-Anya Ma for their support in data preparation.

10. REFERENCES

[1] Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students 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: Springer.Aleven et al., 2006

[2] Aleven, V., & Koedinger, K. R. (2001). Investigations into help seeking and learning with a Cognitive Tutor. In R. Luckin (Ed.), Papers of the *AIED-2001 Workshop on Help Provision and Help Seeking in Interactive Learning Environments* (pp. 47–58).

[3] 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.

[4] Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, *26*(1), 205-223.

[5] Almeda, M. V. Q., Baker, R. S., & Corbett, A. (2017). Help Avoidance: When students should seek help, and the consequences of failing to do so. In *Meeting of the Cognitive Science Society* (Vol. 2428, p. 2433). Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. Cognitive Science, 13(4), 467–505. doi 10.1016/0364-0213(89)90021-9.

[6] Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science*, 13(4), 467–505. doi:10.1016/0364-0213(89)90021-9.

[7] Anderson, J. R. (1993). Rules of the Mind. Hillsdale: Lawrence Erlbaum Associates.

[8] Arroyo, I., Beck, J., Woolf, B. P., Beal, C. R., & Schultz, K. (2000). Macro adapting animal watch to gender and cognitive differences with respect to hint interactivity and symbolism. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), Proceedings of the *5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 574–583). Berlin: Springer Verlag. doi 10.1007/3-540-45108-0_61.

[9] Baker, R. S. J. d., Gowda, S. M., & Corbett, A. T. (2011). Towards predicting future transfer of learning. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), Lecture Notes in Computer Science: Artificial intelligence in Education: *15th International Conference, AIED 2011* (Vol. 6738, pp. 23–30). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-21869-9_6.

[10] Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122.

[11] Beck, J. E., Chang, K., Mostow, J., & Corbett, A. T. (2008). Does help help? Introducing the Bayesian evaluation and assessment methodology. In B. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), Proceedings of the *9th International Conference on Intelligent Tutoring Systems, ITS 2008* (pp. 383–394). Berlin: Springer.

[12] Bong, M., Skaalvik, E. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psych Review*. 15(1), pp. 1-40.

[13] Brown, S. D., & Lent, R. W. (2006). Preparing adolescents to make career decisions: A social cognitive perspective. *Adolescence and Education*, *5*, 201-223.

[14] Childs, D. S. (2017). *Effects of Math Identity and Learning Opportunities on Racial Differences in Math Engagement, Advanced Course-Taking, and STEM Aspiration*. PhD Dissertation. Temple University.

[15] Else-Quest, N. M., Mineo, C. C., & Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity. *Psychology of Women Quarterly*, *37*(3), 293-309.

[16] Goldin, I. M., Koedinger, K. R., & Aleven, V. (2012). Learner differences in hint processing. In K. Yacef, O. Zaïane, A. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), Proceedings of the 5th *International Conference on Educational Data Mining (EDM 2012)* (pp. 73–80). Worcester: International Educational Data Mining Society.

[17] Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European review of social psychology*, *11*(1), 223-255.

[18] Howley, I., Kanda, T., Hayashi, K., & Rosé, C. (2014). Effects of social presence and social role on helpseeking and learning. In G. Sagerer, M. Imai, T. Belpaeme, & A. Thomaz (Eds.), HRI '14: Proceedings of the *2014 ACM/IEEE International Conference on Human-robot Interaction (pp. 415–422)*. New York: ACM. doi:10.1145/2559636.2559667.

[19] Hudley, A. H. C., & Mallinson, C. (2015). *Understanding English language variation in US schools*. Teachers College Press.

[20] Karumbaiah, S., Ocumpaugh, J., Labrum, M., & Baker, R. S. (2019b) Temporally rich features capture variable performance associated with elementary students' lower math self-concept. In *Proceedings of the Workshop on Online Learning and social-Emotional Learning at the 9th International Conference on Learning Analytics and Knowledge* (LAK).

[21] Khachatryan, G. A., Romashov, A. V., Khachatryan, A. R., Gaudino, S. J., Khachatryan, J. M., Guarian, K. R., & Yufa, N. V. (2014). Reasoning Mind Genie 2: An intelligent tutoring system as a vehicle for international transfer of instructional methods in mathematics. *International Journal of Artificial Intelligence in Education*, *24*(3), 333-382.

[22] Kimble, G. A. (1987). The scientific value of undergraduate research participation. *American Psychologist,* 42(3), 267-268.

[23] Klassen, R. M. (2004). Optimism and realism: A review of self-efficacy from a cross-cultural perspective. *International Journal of Psychology*, *39*(3), 205-230.

[24] Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239-264.

[25] Long, Y., & Aleven, V. (2013). Skill diaries: Improve student learning in an intelligent tutoring system with periodic self-assessment. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education, AIED 2013* (pp. 249–258). Berlin Heidelberg: Springer. doi 10.1007/978-3-642-39112-5_26.

[26] Mathews, M., Mitrović, T., & Thomson, D. (2008). Analysing high-level help-seeking behaviour in ITSs. In W. Nejdl, J. Kay, P. Pu, & E. Herder (Eds.), *Adaptive Hypermedia and Adaptive Web-based Systems: 5th*

International Conference, AH 2008 (pp. 312–315). Berlin, Heidelberg: Springer. doi:10.1007/978-3-540- 70987-9_42.

[27] Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., and Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76(2):397–416.

[28] McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human-Computer Interaction*, 5(4), 381–413. doi:10.1207/s15327051hci0504_2.

[29] Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology*, *45*(3), 487-501..

[30] Ogan, A., Walker, E., Baker, R., Rodrigo, M. M. T., Soriano, J. C., & Castro, M. J. (2015). Towards understanding how to assess help-seeking behavior across cultures. *International Journal of Artificial Intelligence in Education*, *25*(2), 229-248.

[31] Pardos, Z. A., & Heffernan, N. T. (2010, June). Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 255-266). Springer, Berlin, Heidelberg.

[32] Razzaq, L., & Heffernan, N. T. (2010). Hints: Is it better to give or wait to be asked? In V. Aleven, J. Kay, & J. Mostow (Eds.), Lecture Notes in Computer Science: Proceedings of the *10th International Conference on Intelligent Tutoring Systems, ITS 2010* (Vol. 1, pp. 115–124). Berlin: Springer.

[33] Roll, I., Baker, R. S. J. D., Aleven, V., & Koedinger, K. R. (2014). On the benefits of seeking (and avoiding) help in online problem-solving environments. *Journal of the Learning Sciences*, 23(4), 537–560. doi:10.1080/10508406.2014.883977.

[34] Schunk, D. H., & Pajares, F. (2005). Competence perceptions and academic functioning. *Handbook of Competence and Motivation*, *85*, 104.

[35] 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.), Proceedings of the *1st International Conference on Educational Data Mining, EDM 2008* (pp. 117–126). Montreal, Canada.

[36] Stamper, J., Barnes, T., & Croy, M. (2011). Enhancing the automatic generation of hints with expert seeding. *International Journal of Artificial Intelligence in Education*, 21(1–2), 153–167. doi:10.3233/JAI-2011-021.

[37] Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613.

[38] Tai, M., Arroyo, I., & Woolf, B. (2013). Teammate relationships improve help-seeking behavior in an intelligent tutoring system. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), Lecture Notes in Computer Science: *Artificial Intelligence in Education* (Vol. 7926, pp. 239–248). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-39112-5_25.

[39] Tessler, R.C., and Schwartz, S.H. 1972. Help seeking, selfesteem, and achievement motivation: an attributional analysis. *Journal of Personality and Social Psychology* 21, 3 (1972), 318-326.

[40] Texas Education Agency. (2018, October 4). *State and School District Summary*. Retrieved from http://www.texaseducationinfo.org/infopage/Summary_Report_Glossary.pdf

[41] Texas Education Agency. (2018, July). *District Type Glossary of Terms*. Retrieved from https://tea.texas.gov/acctres/analyze/1617/gloss1617.html#Major20Urban

[42] The Texas Tribune. (2018, August 24). *State and School District Summary*. Retrieved from https://www.texastribune.org/2018/08/24/texas-school-districts-a-f-grades-takeaways/

[43] Urdan, T., & Pajares, F. (Eds.). (2006). *Self-efficacy beliefs of adolescents*. IAP.

[44] Usher, E. L., & Pajares, F. (2006). Sources of academic and self-regulatory efficacy beliefs of entering middle school students. *Contemporary Educational Psychology*, *31*(2), 125-141.

[45] Vaessen, B. E., Prins, F. J., & Jeuring, J. (2014). University students' achievement goals and help-seeking strategies in an intelligent tutoring system. *Computers & Education*, *72*, 196-208.

[46] VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265.

[47] Wang, Y., & Beck, J. (2013, July). Class vs. student in a bayesian network student model. In *International Conference on Artificial Intelligence in Education* (pp. 151-160). Springer, Berlin, Heidelberg.

[48] Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education*, 33(2/3), 153–169.

[49] Yudelson, M., Fancsali, S., Ritter, S., Berman, S., Nixon, T., & Joshi, A. (2014, July). Better data beats big data. In *Educational Data Mining 2014*.

[50] Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal*, *37*(1), 215-246.

[51] Zeldin, A. L., Britner, S. L., & Pajares, F. (2008). A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *45*(9), 1036-1058.