International Journal of Artificial Intelligence in Education
https://doi.org/10.1007/s40593-021-00272-0

ARTICLE

Context Matters: Differing Implications of Motivation and Help-Seeking in Educational Technology

Shamya Karumbaiah[1] · Jaclyn Ocumpaugh[1] · Ryan S. Baker[1]

Accepted: 26 July 2021

Abstract

Educational technology (EdTech) designers need to ensure population validity as they attempt to meet the individual needs of all students. EdTech researchers often have access to larger and more diverse samples of student data to test replication across broad demographic contexts as compared to either the small-scale experiments or the larger convenience samples often seen in experimental psychology studies of learning. However, the source of typical EdTech data (i.e., online learning systems) and concerns related to student privacy often limit the opportunity to collect demographic variables from individual students—the sample is diverse, but the researcher does not know how that diversity is realized in individual learners. In order to ensure equitable student outcomes, the EdTech community should make greater efforts to develop new methods for addressing this shortcoming. Recent work has sought to address this issue by investigating publicly-available, school-level differences in demographics, which can be useful when individual-level variation may be difficult or impossible to acquire data for. In this study, we use this approach to better understand the role of social factors in students' self-regulated learning and motivation-related behaviors, behaviors whose effectiveness appears to be highly variable between groups. We demonstrate that school-level demographics can be significantly associated with the relationships between students' help-seeking behavior, motivation, and outcomes (math performance and math self-concept). We do so in the context of reasoning mind, an intelligent tutoring system for elementary mathematics. By studying the conditions under which these relationships vary across different demographic contexts, we challenge implicit assumptions of generalizability and provide an evidence-based commentary on future research practices in the EdTech community surrounding how we consider diversity in our field's investigations.

Keywords Fairness · Educational technology · Student demographics · Math learning · Math self-concept · Help-seeking · Motivation

✉ Shamya Karumbaiah
shamya@upenn.edu

Extended author information available on the last page of the article

Published online: 19 August 2021

Introduction

As educational technology (EdTech) researchers and designers seek to support productive learning behaviors, they are faced with a challenge. Complex constructs like motivation, interest, and engagement are known to influence a variety of learning behaviors (Renninger et al., 2018; Ryan & Deci, 2000). However, due to practical constraints of research projects (e.g., budget, recruitment, accessibility, and time), many of these studies involve either small-scale experiments or larger convenience samples of middle-class, undergraduate students (see discussion in Kimble, 1987), which can make it difficult to determine the extent to which these findings will generalize to new and diverse populations of students. Consequently, decisions inspired by such studies could lead to inequitable outcomes for students, if findings are inapplicable for key groups of learners who are not studied. Even when EdTech researchers obtain larger sample sizes, the typical source of EdTech data (i.e., intelligent tutoring systems, adaptive learning platforms, educational game websites) often limits the practicality of obtaining demographic variables from individual students. Beyond practicality (e.g., the ease of acquiring log data on student interactions compared to student demographic data), concerns such as student privacy can reduce the collection of demographic data. 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 (for instance) 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 Childs, 2017) and to constructs related to motivation more specifically (Usher & Pajares, 2006; Zeldin & Pajares, 2000; Zeldin et al., 2008).

Adaptive EdTech with automated decision-making attempts to meet individual student needs, but past research points out that technologies that aim to benefit all students might disproportionately benefit the more advantaged groups. Despite some examples of the success of adaptive EdTech technologies for historically underrepresented groups (i.e., Finkelstein et al., 2013; Huang et al., 2016; Koedinger et al., 1997; Roschelle et al., 2016), learning technologies have not had overall success in closing society's achievement and opportunity gaps (Hansen & Reich, 2015). Despite removing technical and economic barriers (Attewell, 2001), the social and cultural barriers contributing to inequity remain challenging. Institutionalized and unconscious bias and social and cultural distance between EdTech designers and those they seek to serve (especially low-income and minority groups) are the two common sources of failure for the equitable deployment of new technologies (Reich & Ito, 2017). Technology developers' lack of awareness of sociocultural contexts and the needs of different student subgroups can lead to unfortunate consequences. Reich and Ito (2017) emphasize that measuring differences in how various subgroups experience and benefit from EdTech will be a crucial component of our deepening understanding of EdTech and in addressing the inequalities that emerge. At the same time, they acknowledge that there are substantial barriers to collecting relevant demographic data for individual students. Without studies to ensure population validity (Ocumpaugh et al., 2014), such systems may have to compromise on how effectively they can individualize when they are used in unexpected ways by diverse learners (Doroudi & Brunskill, 2019) and in diverse settings. Thus, it seems important that EdTech designers and researchers make greater efforts to overcome the challenges involved in collecting demographic data in order to ensure population validity, equity-oriented EdTech design, and fairness-aware educational data mining (e.g., Ocumpaugh et al., 2014). As such, some researchers have sought to extend student learning models to include information from the broader context, building models at the class, school, or school-cluster level instead of just the student-level (Wang & Beck, 2013; Pardos & Heffernan, 2010, Yudelson et al., 2014).

Our broader research goal is to investigate if the current designs of adaptive EdTech lead to inequitable student outcomes across different demographics. Within this paper, we incorporate broader demographic contexts into an investigation of help-seeking (where a student deliberately asks for assistance in trying to complete or understand a problem) and motivational constructs within EdTech. Help-seeking is chosen as a phenomenon for investigation based on a recent review of the help-seeking literature, which found that the effectiveness of help-seeking behaviors was highly variable across studies (Aleven et al., 2016). Most of the previous research on help-seeking in EdTech has typically used a cognitive lens. Thus, we attempt to shift the focus to also consider social factors. We conduct our investigations in the context of reasoning mind, an intelligent tutoring system for elementary mathematics. Specifically, we demonstrate how readily-available, school-level demographics might reveal how help-seeking and other motivational behaviors of students correlate with two student outcome measures: (1) mathematics performance and (2) mathematics self-concept (an affective measure of students' perception of their own cognitive ability, which is known to predict performance; Lee, 2009). This work has a direct implication to EdTech designers and researchers, who often rely on features such as the universal design of hints and gamification informed by small-scale experiments or larger convenience samples, ignoring group differences.

Prior Work

This section will review the relevant prior work on help-seeking and motivation and the role they play in EdTech design. We also review the past research on the influence of student demographics on help-seeking, motivation, and the two outcome variables—math performance and self-concept.

Previous research on student help-seeking in ITSs has often taken a cognitive approach, focusing on understanding the cognition behind students' choices around help-seeking and the relationship between different forms of help-seeking and student outcomes (Aleven et al., 2016). However, this research has often obtained conflicting findings, including contradictory (positive and negative) correlations between hint usage and learning (Koedinger & Aleven, 2007). Though accounts of these findings have often focused on *how* or *when* students seek help (e.g., Koedinger & Aleven, 2007), we believe that the conflicting findings may also relate to *who* chooses to seek help. In this section, we review prior work on help-seeking, comparing the findings in the ITS research that has typically taken a cognitive approach to research that has explored the social and demographic factors related to help-seeking. We then look at previous research on intrinsic and extrinsic motivation, especially as it relates to social and demographic differences in education, based on evidence that help-seeking is associated with the student's motivation (Butler, 2006; Nelson-Le Gall & Resnick, 1998). Finally, given the relationships between mathematics self-concept and help-seeking (Skaalvik & Skaalvik, 2013), we discuss prior research related to the role that demographics play in math self-concept and math performance, which allow us to better understand how students' behaviors are related both to their actual skill level and their perceptions about that ability.

Given that previous research shows social differences in help-seeking outside of ITS systems and in other constructs related to motivation, we hypothesize that we should expect that help-seeking and motivational behaviors may demonstrate demographic differences. These differences may account for the sometimes contradictory findings that cognitive research has shown when comparing help-seeking practices to student performance.

Help-Seeking

Help-seeking functions—mostly in the form of on-demand, contextual, real-time hints—are common features in most intelligent tutoring systems (ITSs; VanLehn, 2006), and they have long been believed to foster emerging concepts and principles in a student's learning (Anderson, 1993) and to support struggling students during problem-solving (Aleven & Koedinger, 2000). Yet help-seeking behaviors are not always beneficial (Aleven and Koedinger (2000, 2001); Aleven et al., 2016). While much of the prior work on help-seeking in ITSs has focused strictly on its cognitive effects, other research suggests that we should be exploring how motivational social factors may influence these findings, as these patterns may help us to better understand the social issues that govern when and how students choose to engage with help-seeking opportunities.

Help-Seeking: A Cognitive Lens

The literature on help-seeking behaviors in ITSs now stretches back over two decades (see extensive review in Aleven et al., 2016). 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., (1989) 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 Aleven et al., 2016).

After early findings showed a negative correlation between hint usage and student learning in one context (Aleven & Koedinger, 2000), 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) (Aleven et al., 2006). 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 (Aleven et al., 2016; Goldin et al., 2012).

A number of studies have attempted to identify the degree of help needed at any given moment (e.g., Koedinger and Aleven's (2007) assistance dilemma). These studies have shown several interesting findings. For example, (1) on-demand hints lead to greater learning gains than automatic hints in middle-school mathematics (Razzaq & Heffernan, 2010); (2) hint content (goal feedback versus other kinds of feedback) is related to student learning in Geometry (McKendree, 1990); (3) hints about which step to try next to improve student learning of logic proofs (Stamper et al., 2011).

In general, the literature suggests that increasing hint usage does not always lead to better domain-level learning (Aleven et al., 2016). However, the literature on help-seeking in ITSs has produced research that aggregates into a complicated and contradictory narrative, including: (1) a negative association between hint usage and learning (Aleven & Koedinger, 2001); (2) a positive association between hint usage and learning (Beck et al., 2008; Wood & Wood, 1999); (3) a positive association between hint usage and learning only when time per hint level is considered (Long & Aleven, 2013); (4) a positive association between time spent in bottom-out hints and learning (Shih et al., 2008); (5) a negative association between the number of bottom-out hints used and learning (Mathews et al., 2008); (6) positive benefits for students but only when they have a medium level of skill (Roll et al., 2014); (7) a negative association between help avoidance and learning early within practice (Almeda et al., 2017) and on a transfer post-test (Baker et al., 2011). In addition, individual differences in self-regulation were observed in how students process hints and how that impacts their performance (Goldin et al., 2012). Vaessen et al. (2014) found that students' achievement goals (mastery and performance goals) are closely related to their help-seeking and could be used to predict their strategies for help-seeking. Overall, 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.

Help-Seeking: A Social Lens

While the role of social factors on help-seeking behaviors has not been the primary focus of the EdTech community (see Aleven et al., 2016), 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 (Tessler & Schwartz, 1972). Relatedly, Howley et al. (2014) suggest that asking for help (within in-person learning) may trigger experiences of evaluation anxiety—the fear of being judged. In fact, early work on student help-seeking sometimes focused on its maladaptive uses (Baltes, 1997), a categorization that suggests that some students might avoid help after accurately assessing classroom expectations of independence in their learning processes. Meanwhile, Butler (1998) identified three factors related to help-seeking behaviors, including the desire to work autonomously, the desire to demonstrate high ability, and the desire to finish the assignment quickly.

These kinds of concerns seem ripe for sociocultural variation, and a few studies have begun to explore how these differences may emerge. For example, Tai et al. (2013) 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 they 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., Tessler & Schwartz, 1972) and improved student learning.

Student Demographics and Help-Seeking Behaviors

When social expectations guide behaviors, researchers should expect to find demographic differences, and some studies have specifically investigated this with respect to help-seeking behaviors. For example, Ogan et al. (2015) found that the models on effective help-seeking did not transfer well between countries (namely Costa Rica, the Philippines, and the USA). Likewise, Arroyo et al. (2000) 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. This work matches findings in other learning contexts, which has shown both that there may be racial and gendered interactions influencing differences in help-seeking behaviors and that these different behaviors may explain subsequent achievement patterns (Ryan et al., 2009). Combined, these findings suggest that researchers in the ITS community should be paying attention to cultural differences that may influence how students perceive help-seeking opportunities to affect their sense of *competence* and *autonomy*. That is, if we are going to design ways to improve appropriate help-seeking behaviors, we have to understand which students are currently reluctant to use these behaviors.

Intrinsic and Extrinsic Motivation

While help-seeking has been studied in terms of either cognitive or social factors, student motivation tends to be classified into either intrinsic or extrinsic motivational factors. As described in Deci and Ryan's self determination theory (SDT; 1985), "intrinsic motivation refers to doing something because it is inherently interesting or enjoyable and extrinsic motivation refers to doing something because it leads to a separable outcome." There is a general consensus on the role of intrinsic motivation in high-quality learning and creativity, reflecting natural human propensities to learn (Ryan & Deci, 2000). However, the importance of extrinsic motivation is argued to be dependent on autonomy as experienced by the student. Thus, a student's extrinsic motivation could reflect either true self-regulation or external control. Since it is difficult to expect students to be intrinsically motivated by all subject matter or to inherently enjoy all learning activities, educators and EdTech designers often rely on extrinsic motivators. However, passive and controlling forms of extrinsic motivation can leave students only externally propelled into action (e.g., with the expectation of being tested on it). In contrast, more active and volitional forms of extrinsic motivation can win students' acceptance (e.g., with the expectation of teaching it to a peer; see review in Ryan & Deci, 2000).

Social Factors Influencing Intrinsic Motivation

Researches examining the role of autonomy in intrinsic motivation suggest that immediate contextual conditions (e.g., those found in students' schools and homes) can systemically catalyze or undermine the needs of competence and autonomy (Ryan & Stiller, 1991). Cognitive evaluation theory (CET)—a subtheory of SDT—specifies social factors that lead to differences in intrinsic motivation. It argues that interpersonal events and structures that are conducive to feelings of competence and autonomy can elicit, sustain, or enhance intrinsic motivation for the action performed. Examples of such structures include optimal challenge, constructive feedback (Harackiewicz, 1979), and the absence of shaming evaluations (Deci & Cascio, 1972). Along with the increases in perceived competence (Vallerand & Reid, 1984), students must experience their behavior to be self-determined for intrinsic motivation to increase (Ryan, 1982).

In fact, the issue of autonomy versus control has been a popular field of motivation research with considerable controversy. Lepper et al. (1973) first reported that extrinsic rewards could undermine intrinsic motivation. A later meta-analysis argued that any type of expected tangible reward made contingent on task performance undermines intrinsic motivation by shifting the perceived locus of causality from internal to external (Deci et al., 1999). On the other hand, a parallel school of thought has argued against prematurely dismissing the value of tangible extrinsic rewards for students who are not intrinsically motivated (Hidi & Harackiewicz, 2000). Other structures that have been reported to have a negative outcome on intrinsic motivation include deadlines (Amabile et al., 1976), directives (Koestner et al., 1984), and competition pressure (Reeve & Deci, 1996). Students who were overly controlled also learned less and lost their initiative to learn (Grolnick & Ryan, 1987). In contrast, choice and opportunity to engage in self-direction (Zuckerman et al., 1978) were reported to enhance intrinsic motivation. A similar positive effect on intrinsic motivation, curiosity, and desire for challenge was reported with autonomy-supportive teacher practices (Ryan & Grolnick, 1986).

Social Factors Influencing Extrinsic Motivation

An important challenge for educators and EdTech designers is to design activities that, when not intrinsically interesting, could still motivate students to value and self-regulate on their own without external pressure (Zimmerman, 1985). Organismic Integration Theory (OIT)—another sub-theory of SDT—emphasizes the role of student autonomy in designing activities and experiences that improve extrinsic motivation (Ryan & Connell, 1989). Specifically, more autonomous extrinsic motivation is associated with greater engagement (Connell & Wellborn, 1990), better performance (Miserandino, 1996), higher quality learning (Grolnick & Ryan, 1987), and greater psychological well-being (Sheldon & Kasser, 1995).

As with intrinsic motivation, social-contextual conditions that foster a students' feeling of competence and autonomy support self-regulation with extrinsic motivation. In addition, since extrinsically motivated behaviors do not reflect inherent interest, their value to the people, group, or culture whom the student identifies with becomes important (Ryan & Deci, 2000). For example, students' relatedness to teachers in their classrooms, along with their sense of being valued by their teacher, is strongly linked to their adoption of classroom values (Ryan et al., 1994). Similar findings are reported for the importance of autonomy, relatedness, competence supportive practices in extrinsically valued activities (Grolnick & Ryan, 1987; Williams & Deci, 1996).

Student Demographics Influencing Motivation

Demographic differences in the development of students' motivational profiles and a corresponding need for different supports are noted in several studies (Renninger et al., 2018). One notable finding includes a significant linear decrease in intrinsic motivation from 3rd grade through 8th, while extrinsic motivation showed few differences across grade levels (Lepper et al., 2005), a finding which has been replicated in other studies (Gottfried et al., 2001). Similarly, gender has been shown to influence the relationship between students' motivation and the topic or context of the learning task (Hoffmann & Häussler, 1998). Other demographic categories, including those that are more clearly sociocultural (as opposed to biological) in nature, have also proved important to motivation research and interventions. For example, both underrepresented students and first-generation students were positively influenced by interventions involving reflections on utility value or relevance, resulting in increased interest in the subject matter (Hulleman et al., 2016).

Thus, there is a need for more research to look at social factors while studying student motivation and help-seeking in EdTech. Such studies should consider student demographics to understand how to foster positive student outcomes. In this paper, we study students' help-seeking behavior and their intrinsically and extrinsically motivated behaviors in an online math tutor used in traditional classrooms during regular instruction. We aim to shift the focus of EdTech research for constructs like help-seeking from purely cognitive factors to the contextual factors that might play a more prominent role than is assumed. 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 and motivational behaviors.

The Role of Demographics in Predicting Student Outcomes

Two outcome measures are used in this study: math performance and math self-concept. The first helps us to better understand how much help a student might need, while the second helps us to better understand whether how confident they are in their own ability (which may be more linked to help-seeking choices than actual proficiency). 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.

Demographics and Math Performance

The literature addressing demographic differences in learning outcomes 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, many scholars are now instead discussing an opportunity gap, as findings generally show that achievement patterns favor groups for whom the educational system was initially designed (see discussion in Chambers, 2009; Flores, 2007). Scholars point out that reframing this discussion in terms of opportunities to learn emphasizes the need to address the environmental inadequacies that are driving inequitable outcomes (Flores, 2007; Ladson-Billings, 2013). Childs' (2017) analysis shows, for example, that minority students are just as likely to value mathematics as other students but are less likely to attend schools where advanced mathematics classes are offered.

However, less tangible cultural and linguistic differences may also play a role. We know, for example, that the strategies for speech act like asking questions can vary substantially even in the same language. (See, for example, Greenbaum & Greenbaum's (1983) review of classroom practices among different Native American groups or Chavajay and Rogoff's (2002) review of the literature on classroom practices among cultures that do not use known-answer questions.) If students' patterns of communication are different from those expected by educators, their attempts at communication—including help-seeking—may not receive appropriate responses (Hudley & Mallinson, 2015). 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 if the help-seeking system were appropriately designed.

Demographics and Self-Concept

Demographic variables have also been shown to correlate with constructs like math self-concept (self-beliefs related to a specific task; Bandura, 1982). Math self-concept (sometimes used interchangeably with self-efficacy, although see Bong and Skaalvik (2003) for discussion) has been found to be a predictor of various measures of achievement and career choice (see Brown & Lent, 2006), although the relationship between self-concept and mathematics achievement does vary in magnitude in different countries (Wilkins, 2004). It has also been linked to motivational constructs, including achievement goal orientation, anxiety, and self-concept (Schunk & Pajares, 2005).

Early work proposed that self-efficacy was a product of a person's own accomplishments and the feedback they receive on their work (Bandura, 1982; Urdan & Pajares, 2006); however, more recent studies have indicated that the source of self-efficacy may vary along demographic lines like gender and ethnicity (Usher & Pajares, 2006; Zeldin & Pajares, 2000; Zeldin et al., 2008). For example, Klassen's (2004) investigation of self-efficacy among seventh-grade students found that ethnic majority students followed Bandura's (1982) predictions, citing personal achievements as a source of self-efficacy, but ethnic minority students were more likely to cite group capabilities for collective efficacy. By contrast, Else-Quest et al. (2013) studied the intersection of gender, ethnicity, and achievement in tenth-grade students from a large northeastern city and found that males reported higher math self-concept and expectation of success as compared to 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 interactions (Zeldin et al., 2008), 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 (Steele, 1997). Previous research shows that assimilation to social identity (e.g., gender and cultural identity) increases when people are experiencing uncertainty (Hogg, 2000). This could suggest that students could become more susceptible to negative cultural stereotypes (e.g., Steele, 1997), particularly those related to STEM performance, during periods of confusion associated with learning, making help-seeking an important behavior to study.

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.

Data Collection

Reasoning Mind

This study analyzes data from students using Imagine Learning's Reasoning Mind (RM) foundations (Fig. 1), an intelligent tutoring system for elementary mathematics. The majority of Reasoning Mind's students are in Texas, but they represent a range of traditionally underrepresented populations across rural, urban, and suburban schools. RM includes features that are designed to mimic other social experiences in the classroom, including both virtual peers and the system's signature pedagogical agent, known as the Genie, that guides students in their learning.

In this blended environment, students learn through self-paced problem solving, interactive explanations, and skill-based games. Problem sets are classified into three

Fig. 1 Left—Reasoning Mind Foundations home screen; Right – An example problem displaying the Genie

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. Reasoning Mind Foundations is generally used in a classroom environment. 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 (see review in Khachatryan et al., 2014).

Hints in Reasoning Mind

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

Intrinsic and Extrinsic Motivations in Reasoning Mind

Most studies on motivation rely on student self-reports, which are dependent on how self-aware and reflective participants are (Renninger et al., 2018). Instead, we use student choices in the online math tutor as proxies for their intrinsic or extrinsic reason for it (cf. Barron et al., 2014). When students practice their skills in RM Foundations, they are awarded points for solving problems correctly (and more points if they are consistent). They can use these points to buy virtual prizes or items like e-books, animations, and decorations for a virtual room called "My Place" (Fig. 3), a feature seen in other learning systems as well. These extrinsic motivators are analogous to the game-like features in other EdTech systems such as iSTART-ME, the Motivational Enhanced (ME) version of iSTART, an intelligent tutoring system for reading (i.e., Jackson et al., 2009).

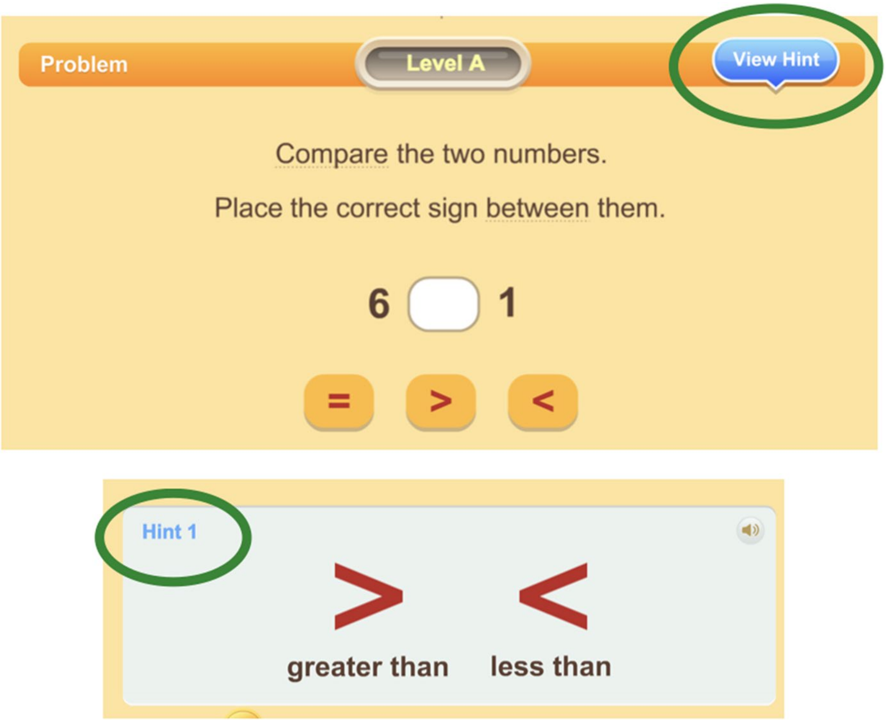

Fig. 2 Top—Problem screen with a button to view hint (highlighted in green); Bottom—Hint displayed to the student when they request to view

Fig. 3 Left – "My place" lobby with entrances to library and great hall. At the bottom are the points awarded; Middle – Library with books and movies purchased using points; Right – The great hall decorated with furnishing items purchased using points

Another automatic outcome of earning more points and streaks in RM Foundations is that doing so opens up more challenging and optional problems (B and C level), which are otherwise locked until students demonstrate mastery in simpler problems (A level). Since B/C level problems are almost always optional, a student can choose to continue working on A level problems, allowing them to more easily earn points to purchase more items. The desire for challenge within the task rather than external rewards is a strong indicator of intrinsic motivation (Ryan & Deci, 2000), as is the will to pursue the learning task when it is a "free choice" (Deci, 1971). Thus, we use the number of items purchased and the number of B and C level problems attempted as proxies for students' extrinsic and intrinsic motivation, respectively. Our proxy for differentiating extrinsic motivation from intrinsic assumes that students understand that it is easier to earn points by solving the simpler problems than it is to do so by taking on more challenging material.

Participant Schools

We analyze 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, in schools where at least 25 students were using the software. There is a total of 9122 2nd through 5th-grade students in this data (4749 2nd graders, 1964 3rd graders, 1582 4th graders, and 827 5th graders)—i.e., Reasoning Mind was more widely used in 2nd-grade classes than older students. However, there was considerable variation in the use of Reasoning Mind across grades in different schools – the standard deviations of the proportion of grades across schools are 33.06%, 16.87%, 15.03%, and 19.29% for grade 2, grade 3, grade 4, and grade 5 respectively. 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 3039 students across 62 schools.

Comprehensive log data captured student interactions with the system for the entire period, resulting in data for all 9122 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 2238 students in 22 schools.

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.

Exploring Help-Seeking

From the interaction log data, we operationalize help-seeking behavior as the number of hints used by a student in Reasoning Mind Foundations. As shown in Fig. 4 (left), students in this study averaged less than 30 hint requests annually (mean = 27.01, SD = 55.72). The overall low hint usage could be attributed to RM being highly scaffolded—anticipating many student questions beforehand.

Exploring Intrinsic Motivation

As described in Sect. 3.3, we use the student choice of solving advanced and optional B and C-level problems as our proxy for intrinsic motivation. Accordingly, we

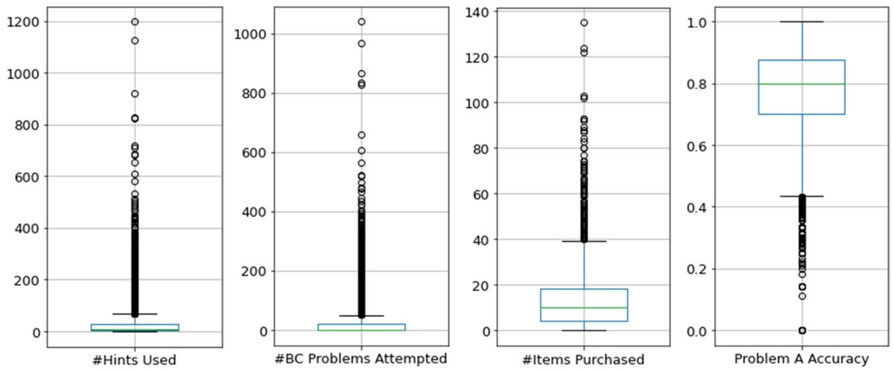

Fig. 4 From left to right: Distribution of the number of hints (leftmost), number of B and C-level problems attempted, number of items purchased, and math performance (accuracy in A-level problems; rightmost). The middle line in the box indicates the median value

operationalize intrinsically motivated behavior as the number of B and C-level problems attempted by a student in Reasoning Mind Foundations. As shown in Fig. 4 (left), students in this study averaged less than 30 B and C-level problems annually (mean = 26.81, SD = 61.60).

Exploring Extrinsic Motivation

As described in Sect. 3.3, we use the student choice of buying items (virtual prizes) to decorate a virtual room called "My Place" as our proxy for extrinsic motivation. Accordingly, we operationalize extrinsically motivated behavior as the number of items purchased by a student in Reasoning Mind Foundations. As shown in Fig. 4 (left), students in this study averaged less than 15 item purchases annually (mean = 12.51, SD = 10.87).

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, i.e., the ratio of the number of correct answers to the number of problems attempted. We choose only A-level problems because they represent the core curriculum within the software. We obtain the problems attempted and the correctness of student answers from the interaction log data. As presented in Fig. 4 (right), student-level calculations show a mean of 0.77 (SD = 0.14)—i.e., students obtained correct answers 77% of the time.

Exploring Math Self-Concept

Students' self-concept in mathematics was measured using a five-item survey adapted from Marsh et al. (2005). 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" and "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. Our previous work used this data to build a predictive model of math identity-related constructs like self-concept using language and behavior patterns (Crossley et al., 2020).

The distribution of students' responses is given in Fig. 5 (self-concept pre: mean = 2.64 standard deviation = 0.77; self-concept post: mean = 2.44, standard deviation = 0.80). As summarized in Marsh et al. (2005), domain-specific self-concept (e.g., mathematics self-concept) shows developmental patterns of decline from early childhood to adolescence and then increases during early adulthood. We see a similar pattern in our student population, with the self-concept post-test score statistically significantly lower than the pre-test ($t = 5.2$, $p < 0.001$). The internal consistency of these items was found to be satisfactory, with a Cronbach's $\propto$ of 0.74.

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 Figs. 6, 7, and Table 1, there is considerable variance in the variable aggregates (mean) across the schools, especially in hint usage and math performance.

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 some contextual factors that are likely to affect the school culture or climate and thereby may affect student use of RM Foundations.

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) Limited English Proficiency

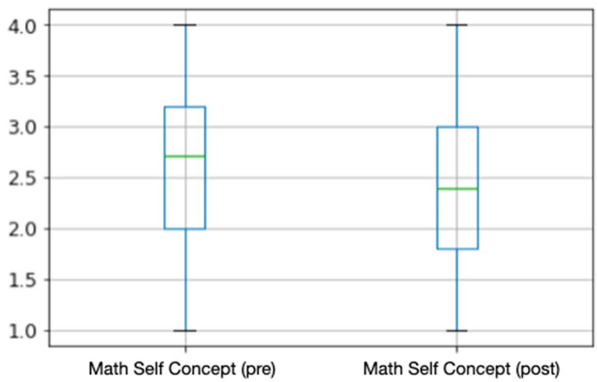

Fig. 5 Distribution of the pre and post measures of math self-concept

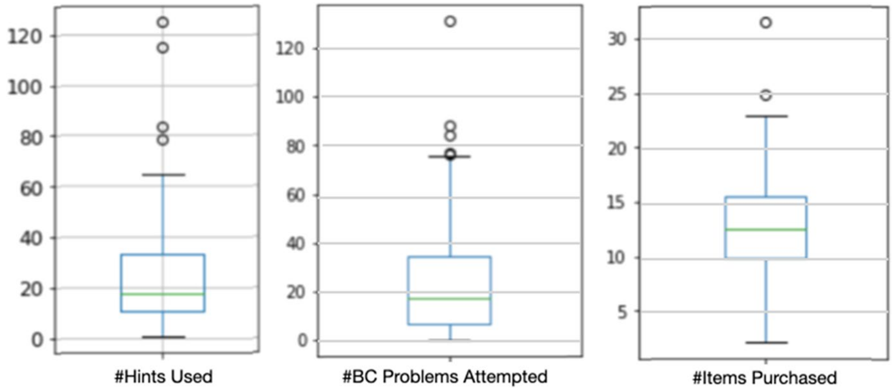

Fig. 6 Distribution of school-level aggregates of the variables

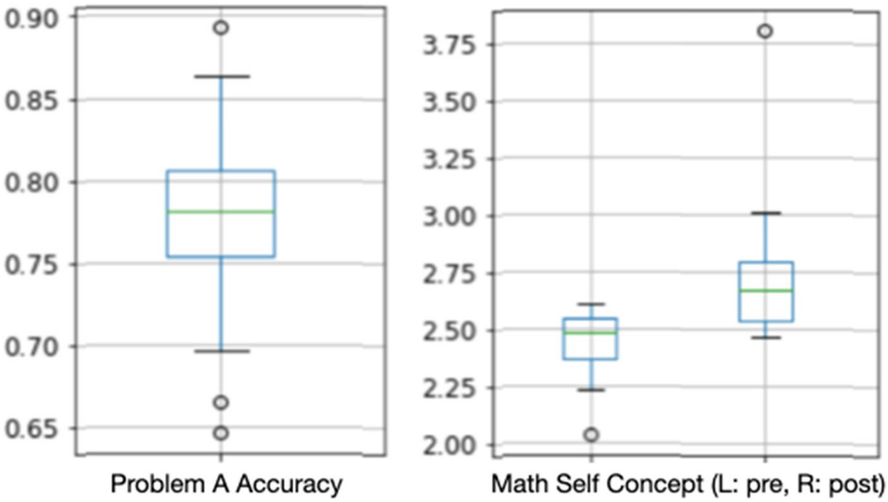

Fig. 7 Distribution of school-level aggregates of the outcomes

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

	Mean	SD
Hint Usage	24.52	21.30
Number of B & C level problems attempted (Proxy for intrinsic motivation)	23.79	24.04
Number of Items purchased (Proxy for extrinsic motivation)	12.82	4.91
Math performance	0.78	0.04
Math self-concept (pre)	2.69	0.35
Math self-concept (post)	2.43	0.15

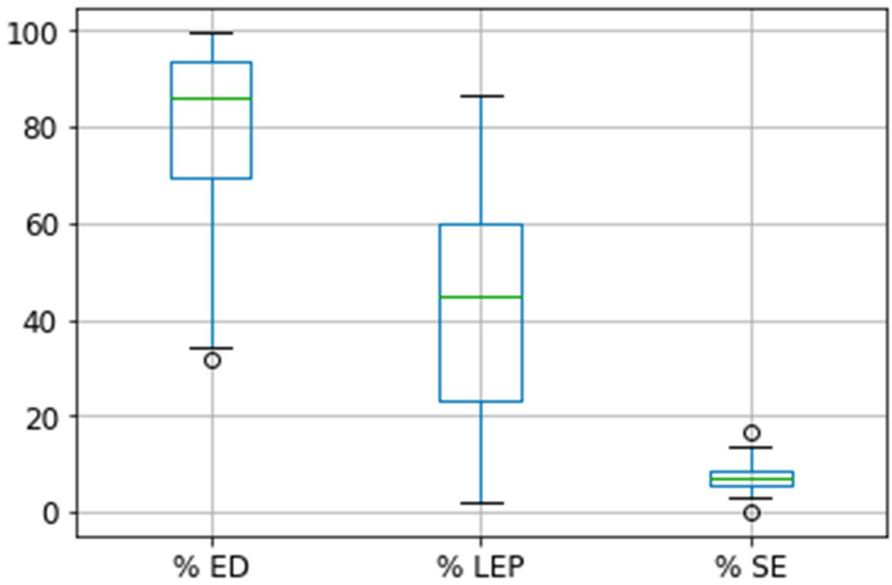

Fig. 8 Distribution of non-binary school-level demographics for the 110 schools selected in this study. *ED* economically disadvantaged; *LEP* limited English proficiency; *SE* special education

Table 2 Mean and standard deviation (SD) of the school-level demographics

	Mean	SD
% Economic disadvantage	78.3	16.6
% Limited english proficiency	41.4	20.6
% Special education	6.9	3.1
Urbanicity (binary)	60.4%	-
Charter (binary)	27.1%	-

(LEP), or as (3) Special Ed (SpEd), as well as (4) the urbanicity of the school and whether or not it is a (5) charter school. These terms are defined by the State of Texas as follows (TEA, 2018a). Students are classified as EcD if they qualify for free or reduced-price meals under the National School Lunch and Child Nutrition Program; it is worth noting that a large proportion (avg = 40%) of Texas public school students qualify for this status (TEA, 2018a). 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 (2018b) classifies a school district as urban (or not) based on whether its school district (a) is located in a county with a population of at least 960,000; and (b) has the largest enrollment in the county or

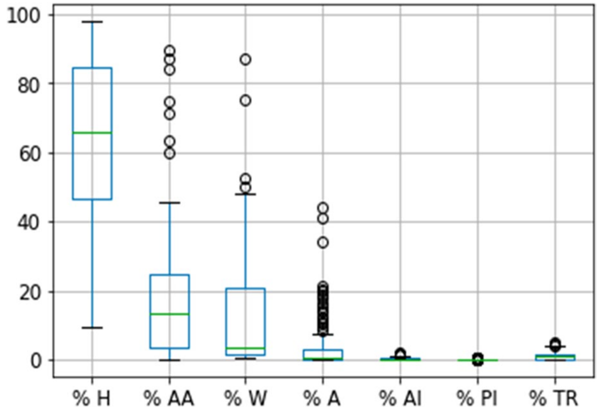

Fig. 9 Distribution of percentages of school-level ethnicities for the 110 schools selected in this study. *H* Hispanic; *AA* African American; *W* White; *A* Asian; *AI* American Indian; *PI* Pacific Islander; *TR* two or more races

Table 3 Mean and standard deviation (SD) of the school-level percentages of ethnicities

	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

*Categories constituting less than 5% of the data were excluded from further analysis

its enrollment is greater or equal to 70% of county's largest district. As seen in Table 2 and Fig. 8, 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, using the categories provided by the TEA. As Table 3 shows, Hispanic students (the TEA's term) 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 Fig. 9the 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.

Analysis

Our data exploration (Sect. 4) suggests that help-seeking behavior, intrinsic and extrinsic motivation, math performance, math self-concept, and demographics each vary by school. Our goal with the analysis is to investigate whether the relationship between the behavior (hint usage and motivational behaviors) and the outcomes vary for different student populations. Thus, we conduct a two-step data analysis to explore how help-seeking and the two motivations might differ based on student demographics, while controlling for performance and math self-concept.

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

Results

Relationship Between Variables and Student Outcomes

Help-Seeking and Student Outcomes

Figure 10 shows 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)).

Help-Seeking and Math Performance

Grouping students by schools allows us to see that the relationship between hint usage and math performance differs in important ways, even before we look at

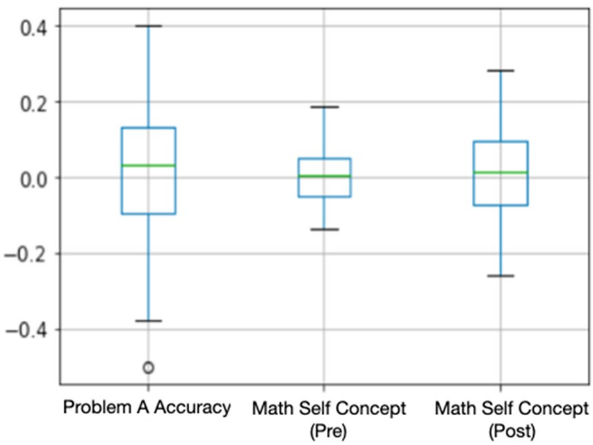

Fig. 10 School-level correlations between hint usage and math performance versus the correlations between hint usage and self-concept measures

demographic variables more directly. Specifically, when student measures are aggregated at the school level, as they are in Fig. 10, the correlation between hint usage and math performance ranges from − 0.39 to 0.40 (SD = 0.18). In contrast, when we do not aggregate students into school-level populations (instead, treat them all as a single population), there is not a significant relationship between hint usage and math performance (ρ = − 0.008, p = 0.44). In other words, while there is an appearance of varied effects within individual schools, it appears to cancel out when considering all schools together.

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, as shown in Fig. 10, 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 essentially zero (ρ = − 0.008, p = 0.442 for pre and ρ = − 0.007, p = 0.77 for post).

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 = 0.084 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 that hints are associated with positive outcomes in some student populations but not in others.

As seen in Fig. 10, the median of the correlations is centered close to zero. For these schools, there is no association between hint usage on student outcomes. Figure 10 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.

Motivational Behaviors and Student Outcomes

Figure 11 shows the distribution of correlations across schools between the students' motivational behaviors (number of B and C level problems attempted and number of items purchased) and their outcomes (math performance and the pre- and post-measures of math self-concept).

Intrinsic Motivation, Extrinsic Motivation and Math Performance

When student measures are aggregated at the school level, we see that the correlation between the intrinsic motivation behaviors and math performance ranges from

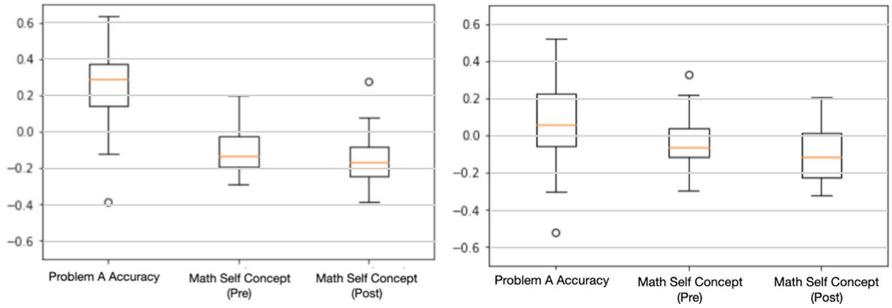

Fig. 11 Distribution of correlations across schools between student outcomes and the proxies of their intrinsic (left) and extrinsic (right) motivation

− 0.12 to 0.62 (SD = 0.18). The majority of the schools show a significantly positive correlation between intrinsic motivation and performance, in line with the vast body of empirical research confirming this relationship (Ryan & Deci, 2000). When we do not aggregate students into school-level populations (instead treat them all as a single population), there is still a significant relationship between intrinsic motivation and math performance ($\rho = 0.17$, $p < 0.001$) with a correlation value closer to the average across schools but doing so fails to capture the range and variance in this relationship.

We also find that the correlation between extrinsic motivation behaviors and math performance ranges from − 0.31 to 0.52 (SD = 0.19). When we do not aggregate students into school-level populations, there is still a significant relationship between extrinsic motivation and math performance ($\rho = 0.04, < 0.001$), but the correlation value is very close to zero.

The school-level correlations between the motivations and math performance range from − 0.12 to 0.62 for intrinsic motivation and – 0.31 to 0.52 for extrinsic motivation (SD = 0.18 for intrinsic motivation and SD = 0.19 for extrinsic motivation). This variance indicates that the relationship between the motivations and student performance differs by student populations. The average correlation, when grouping students by schools and taking an average across the schools, is 0.26 for the intrinsic motivation, while extrinsic motivation has a relatively lower average correlation of 0.08. The range of the correlations is also shifted upwards for intrinsic motivation when compared to the extrinsic motivation—with a tail of − 0.12 vs. − 0.31, head of 0.62 vs. 0.52, and a similar standard deviation. Our results show that there are school-level differences in the correlation between the different motivations and student performance, with extrinsic motivation showing more negative associations in some schools than others.

Intrinsic Motivation, Extrinsic Motivation 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 the motivations and math self-concept show a relatively wide range, although not as prominent as with math performance.

For pre-year surveys, the correlations for intrinsic motivation ranges from − 0.27 (students with lower self-concept are most likely intrinsically motivated) to 0.21 (students with higher self-concept are most likely intrinsically motivated), shifted lower for post-year survey correlations (− 0.38 to 0.15). When the students in this data were treated as a single population, the correlations are close to zero, though statistically significantly for pre ($\rho = -0.067$, $p = 0.002$) and statistically significantly negative for post ($\rho = -0.163, < 0.001$).

For pre-year surveys, the correlations for extrinsic motivation ranges from − 0.29 to 0.23, which is similar to the post-year survey correlations (− 0.33 to 0.21). When the students in this data were treated as a single population, the correlations are very small or close to zero for both pre ($\rho = -0.037$, $p = 0.12$) and post ($\rho = -0.099$, $p < 0.001$).

The schools at the tail ends of these distributions are interesting case studies. They represent the cases where hint usage and motivations have either a notably high positive correlation or a notably high negative correlation with our outcome measures. 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.

The Influence of School Demographics

Demographic Differences in the Relationship between Help-Seeking and Student Outcomes

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

Table 4 Correlations between school-level demographics and the correlations resulted between student outcomes (math performance, self-concept) and help-seeking

	Correlation between number of hints and		
	Math performance	Self-concept Pre	Self-concept Post
Urbanicity (binary)	**0.292** (0.002)	0.130 (0.564)	0.080 (0.729)
% Economic Disadvantage	**0.256** (0.007)	0.182 (0.417)	− 0.288 (0.205)
% Limited English Proficiency	**0.314** (0.001)	**− 0.452** (0.035)	**− 0.565** (0.008)
% Special Education	− 0.002 (0.982)	**0.463** (0.030)	**0.444** (0.044)
Charter Status (binary)	− 0.083 (0.389)	− 0.058 (0.799)	− 0.269 (0.225)

p-value in parenthesis

Significant correlations after Benjamini and Hochberg post hoc corrections in bold

Table 5 Correlations between school-level ethnicity and the correlations resulted between student outcomes (math performance, self-concept) and help-seeking

	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)

p-value in parenthesis

No significant correlations were obtained after Benjamini and Hochberg post hoc correction was conducted

performance and math self-concept). These findings are summarized in Tables 4 and 5.

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 ($\rho = 0.292$, $p = 0.002$) as well as differences in the percentage of students who are economically disadvantaged (EcD; $\rho = 0.256$, $p = 0.007$) and limited English proficiency (LEP; $\rho = 0.314$, $p = 0.001$). Specifically, more hints are associated with higher math performance among urban students, but more hints are associated with lower math performance among suburban/rural students. In schools with a higher percentage of students who are economically disadvantaged (EcD) or limited English proficiency (LEP), the use of hints is associated with higher math performance. However, as Table 5 shows, other demographic categories that are often considered in educational research, namely ethnicity/race, are not predictive in this context.

School-level Demographics, Help-Seeking, 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 (LEP; $\rho = -0.452$, $p = 0.035$ for pre; $\rho = -0.565$, $p = 0.008$ for post), and the percentage of students in special education (SpEd; $\rho = 0.463$, $p = 0.030$ for pre; $\rho = 0.444$, $p = 0.044$ for post). Specifically, in schools that serve a higher percentage of LEP students, students with low self-concept are more likely to use hints, while in schools with fewer LEP students, students with high self-concept are more likely to use hints. 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, their students with high self-concept use more hints, whereas that relationship is negative in schools that serve fewer SpEd students. This finding is consistent across the start of the year and end of the year surveys.

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

Group Differences in the Role of Motivations in Student Outcomes

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 6 and 7.

School-level Demographics, Intrinsic & Extrinsic Motivations, and Math Performance

As Table 6 (above; column 2) shows, the relationships between the intrinsic motivation and math performance differ significantly in terms of the school's urbanicity ($\rho = 0.281$, $p = 0.003$), whether or not it is charter ($\rho = -0.244$, $p = 0.012$) as well as differences in the percentage of students who are economically disadvantaged (EcD; $\rho = 0.230$, $p = 0.018$) and in special education (SpEd; $\rho = 0.247$, $p = 0.011$). Specifically, there is a positive association between higher usage of more advanced (B- and C-Level) problems and math performance among students from urban schools, and a negative association between higher usage of more advanced (B- and C-Level) problems and math performance among students from suburban/rural schools. More attempts of advanced problems are associated with higher math performance among non-charter students, but more attempts of advanced problems are associated with lower math performance among charter students. Schools with a higher percentage of students who are economically disadvantaged (EcD) have a positive association between higher usage of more advanced problems and math performance. Finally, schools with a higher percentage of students in special education (SpEd) have a positive association between higher usage of more advanced problems and math performance.

In contrast, Table 6 (above; column 3) shows that the relationships between extrinsic motivation and math performance differ significantly in terms of the school's urbanicity ($\rho = -0.238$, $p = 0.010$), whether or not it is charter ($\rho = 0.217$, $p = 0.022$) as well as differences in the percentage of students who have limited English proficiency (LEP; $\rho = -0.226$, $p = 0.017$). Specifically, the association between higher usage of the gamified extrinsic motivations and math performance is positive among students from suburban and rural schools and negative among students from urban schools. Furthermore, the association between

Table 6 Correlations between school-level demographics and the correlations resulted between student outcomes (math performance, self-concept) and behaviors related to intrinsic and extrinsic motivation

	Correlation between behaviors related to motivation and					
	Math performance		Self-concept Pre		Self-concept Post	
	Intrinsic Motivation Behavior	Extrinsic Motivation Behavior	Intrinsic Motivation Behavior	Extrinsic Motivation Behavior	Intrinsic Motivation Behavior	Extrinsic Motivation Behavior
Urbanicity (Binary)	**0.281** (0.003)	**− 0.238** (0.010)	0.186 (0.408)	0.037 (0.869)	0.132 (0.510)	− 0.049 (0.826)
% Economic Disadvantage	**0.230** (0.018)	− 0.151 (0.116)	0.021 (0.926)	0.280 (0.205)	0.275 (0.227)	**0.499** (0.017)
% Limited English Proficiency	0.161 (0.100)	**− 0.226** (0.017)	0.038 (0.863)	0.024 (0.914)	0.018 (0.937)	0.369 (0.090)
% Special Education	**0.247** (0.011)	0.0548 (0.569)	0.167 (0.459)	0.117 (0.604)	0.149 (0.519)	0.109 (0.627)
Charter (Binary)	**− 0.244** (0.012)	**0.217** (0.022)	0.086 (0.702)	0.072 (0.750)	0.167 (0.469)	0.115 (0.610)

p-value in parenthesis

. Significant correlations after Benjamini and Hochberg post hoc corrections in bold

Table 7 Correlations between school-level ethnicity and the correlations resulted between student outcomes (math performance, self-concept) and behaviors related to intrinsic and extrinsic motivation

	Correlation between behaviors related to motivation and					
	Math performance		Self-concept Pre		Self-concept Post	
	Intrinsic Motivation Behavior	Extrinsic Motivation Behavior	Intrinsic Motivation Behavior	Extrinsic Motivation Behavior	Intrinsic Motivation Behavior	Extrinsic Motivation Behavior
% Hispanic	0.012 (0.903)	− 0.203 (0.034)	0.065 (0.774)	0.196 (0.382)	0.116 (0.616)	0.427 (0.047)
% African American	0.072 (0.463)	− 0.038 (0.696)	0.045 (0.844)	− 0.290 (0.190)	0.173 (0.454)	− 0.161 (0.474)
% White	− 0.091 (0.356)	0.207 (0.030)	− 0.408 (0.056)	− 0.191 (0.395)	− 0.192 (0.404)	− 0.123 (0.585)
% Asian	− 0.203 (0.038)	0.081 (0.398)	− 0.045 (0.842)	− 0.330 (0.134)	− 0.008 (0.971)	− 0.312 (0.158)

p-value in parenthesis

No significant correlations were obtained after Benjamini and Hochberg post hoc correction was conducted higher usage of the gamified, extrinsic motivation behaviors, and math performance is positive among students from charter schools and negative among students from non-charter schools. Schools with a higher percentage of students who have limited English proficiency (LEP) show a negative relationship between extrinsic motivation and math performance. Also, like in the hint usage results, ethnicity/race are not significantly correlated in this context as well (Table 7). As such, the correlations between extrinsic motivation and math performance across schools range relatively more negative than intrinsic motivation (− 0.12 (intrinsic) vs. − 0.31 (extrinsic); Sect. 6.1.2), suggesting the behaviors associated with extrinsic motivation may be more harmful to certain schools.

An interesting pattern in these results is the inverse relationship of the same demographic variable with intrinsic and extrinsic motivation. For example, suburban and rural schools have a negative association between math performance and intrinsic motivation and a positive association between math performance and extrinsic motivation. This result suggests that more support may be needed to transition students in rural and suburban settings from gamified extrinsic motivators to develop competence in more advanced and challenging math problems. Similar support may also be needed for students in charter schools. Another concerning trend among schools with a higher percentage of students who have limited English proficiency is the negative association between extrinsic motivation and math performance and the lack of a strong relationship between intrinsic motivation and math performance. This could imply that the gamified motivation in the system may not be enough to improve math competence in these schools, while their students also fail to find motivation in the increased challenge in advanced problems. On the other hand, schools with a higher percentage of economically disadvantaged students could be inherently motivated to solve challenging problems, based on the correlation between the percentage of students economically disadvantaged and intrinsic motivation and the lack of correlation between economic disadvantage and extrinsic motivation. The same is true for schools with a higher percentage of special education students, though intrinsically motivated behaviors may be promoted in this case by the teachers' aides and specialists helping these students.

School-level Demographics, Intrinsic & Extrinsic Motivations, and Math Self-Concept

The relationships between the usage of extrinsic motivation and math self-concept (post) differ significantly in terms of the percentage of students who are economically disadvantaged ($\rho = 0.499$, $p = 0.017$). This relationship is not consistent across the start of the year and end of the year surveys, unlike what we observed in hint usage (Table 4). Also, other demographic factors from Table 6 that were predictive of the relationship between the motivations and math performance were not significant for the relationship between the motivations and math self-concept, neither were the ethnicity/race variables (Table 7).

Discussion

Talking about the social and ethical impacts of computer bias, Garcia (2016) said, "the side effects of unintentionally discriminatory algorithms can be dramatic and harmful." In recent years, data-driven systems have come under scrutiny for amplifying existing social inequities or, in some cases creating new ones (Garcia, 2016). As such, there has been increasing amounts of research on fairness in data and machine learning systems. However, much of this work has focused on optimizing systems based on abstract universal notions of fairness or de-contextualized quantitative metrics and ignoring social, political, and cultural deliberation (Green & Hu, 2018). However, education is a context where achieving fairness with sociotechnical solutions poses unique challenges tied closely to the sociocultural aspects of the domain (Ocumpaugh et al., 2015; Ito, 2017; Karumbaiah et al., 2019, 2021). Despite the increase in the use of data analytics and AI-based systems in education, relatively little work has focused on establishing what fairness means in this context and exploring approaches to achieving it (Holstein & Doroudi, 2019; Slade & Prinsloo, 2013; Subotzky & Prinsloo, 2011). The current literature on ethics in the field has mostly been interested in issues of data ownership and privacy and institutional and policy level considerations (Draschler et al., 2015; Tsai & Gasevic, 2017). Through this paper, we hope to contribute to the emerging conversations on fairness and equity in EdTech systems in two ways—(1) by investigating inequitable outcomes across student subpopulations on two of the fundamental psychological constructs in EdTech design, (2) by demonstrating the use of publicly-available, school-level demographics for fairness research where individual student demographics may be difficult or impossible to acquire.

Difference in Outcomes Across Student Subpopulations

Help-Seeking

Hint-seeking behaviors have been a source of interest among EdTech researchers since the early days of the adaptive EdTech field (Aleven & Koedinger, 2000; Anderson, 1993), yet understanding which hints are effective, for whom, and under what conditions, remains a somewhat elusive goal. 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 hints to improve their understanding of the material, but as these results show, students who are struggling do not always make use of available resources effectively. Within this data (with a relatively low hint usage overall)—which involves students in the same state using the same mathematics learning system—there are also schools where hint usage is associated with low-performance. 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 ineffective at helping these particular students to learn the material.

At least part of the school-level differences in the correlation between hint usage and math performance 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 do not appear to be benefitting from hint use. Schools with fewer students receiving free or reduced-price lunch are more likely to have low performers who do not appear to be benefitting from hint use. Schools in large urban centers are less likely to have low-performing students who do not appear to be benefitting from hint use.

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 with low math self-concept appear to use more hints (in those schools). However, in schools that serve more SpED students, the relationship between self-concept and hint usage is positive (i.e., students who are relatively more sure of themselves ask for more hints). It is also possible that the smaller number of schools 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, such as dividing LEP students based on different languages. However, it is also possible that some of the linguistic differences that influence classroom practices of different ethnic groups (e.g., Hudley & Mallinson, 2015)—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 a button to request a hint.

Intrinsic and Extrinsic Motivations

Research on fostering student motivation has a long history of research in psychology. There have been mixed opinions on the efficacy of extrinsic motivations like tangible rewards, while there is more consensus on the important role of intrinsic motivation in high-quality learning (Deci & Ryan, 2000). Given that many students are not intrinsically motivated in any given subject, EdTech designers often turn to features that can increase extrinsic motivation, such as rewards and gamified activities. Research suggests that the design of extrinsic motivators should value student autonomy and foster self-regulation instead of exerting external control (Ryan & Connell, 1989). Our literature review establishes the role of social factors and the need for different supports for different learners to catalyze student competence and autonomy. However, very little research in EdTech has explored the role of social context in the efficacy of extrinsic motivation.

In this data, we observe that the relationship between students' intrinsic and extrinsic motivation and their outcomes (math performance, math self-concept) is associated with student demographics. In the case of our binary variables, urbanicity and charters, we find the inverse correlations between math performance and intrinsic versus extrinsic motivations. Suburban and rural schools (like charter schools) have a negative association between math performance and intrinsic motivation and a positive association between math performance and extrinsic motivation. Our results suggest that students from rural and suburban schools may need better supports to transition from gamified extrinsic motivations to developing interest in more advanced and challenging math problems. Similarly, our results suggest that such supports may also be needed for students in charter schools. A rather concerning trend in schools with a higher percentage of students who have limited English proficiency is the negative relationship between extrinsic motivation and students' math competence. These results suggest that EdTech designers need to pay special attention to their students' demographic context when designing extrinsic motivations in EdTech systems if our goal is to deliver equitable student outcomes. Overall, if future studies establish a causal link, then we recommend that the system behavior should change based on the student needs. For instance, the design of extrinsic rewards and hints should vary based on student self-concept. At a minimum, teachers should have the ability to override system decisions or make changes to certain design elements based on their assessment of varying student needs, to address possible mismatches between design and the needs of specific learner populations.

Implications for EdTech Designers and Researchers

One of the main implications of this paper for EdTech designers is that a universal design that focuses on improving student outcomes while ignoring individual or group differences might not produce the desired results. One key finding of this paper is that the group differences that matter most for design might not be the groups that are the most immediately obvious. The work presented here is a step towards understanding which group differences may matter, but as recommended by Baker et al. (2019), research in this area should explore a broader range of conceptualizations of context and identity than are currently considered. Ultimately, the vision of culturally aware tutoring systems (Blanchard & Mizoguchi, 2008) can only be fully achieved if we know which groups to adapt them to, and how. Thus far, however, personalization in help design has not taken these types of issues into consideration, primarily focusing on understanding student cognition to provide hints based on the pedagogical content (Aleven et al., 2016). Our findings suggest that developers and researchers on adaptive EdTech should explore broader contextual factors to analyze the effectiveness of hint usage across student subpopulations, and adapt help to a broader vision of student need. Similarly, features to increase extrinsic motivation, such as gamification, may not always be beneficial to academically unmotivated students, depending on group differences.

To illustrate this, let's take the example of students with limited English proficiency. As shown in Sect. 6.2, there is an inverse relationship between 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 for ITS designers to investigate further. Is the text-heavy nature of the hints contributing to this finding? Are limited English proficient students using hints to improve on their math skills, but 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 choose which representation to use? In summary, including school-level demographics to the analysis of complex constructs like help-seeking is an important step in developing designs that are appropriate for all learners.

Our results suggest that research on complex yet widely-used constructs related to student learning and engagement may not generalize well across diverse student populations, especially when the studies are conducted in a small-scale or with convenience samples. Although this finding is not novel in itself, this paper demonstrates an approach to assess the generalizability of the EdTech research findings by using publicly-available, school-level demographics when we have access to larger data (e.g., interaction logs), which may be particularly important when access to individual student demographics is restricted. A broader implication of our work for EdTech researchers is to consider the student demographic factors when explaining contradictory findings on the relationship between student behaviors and outcomes in virtual learning environments. As such, we would echo Paquette et al.'s (2020) recommendation that the research community pay more attention to student demographics, including both commonly reported categories of gender and race/ethnicity, and factors like LEP, EcD, SpEd, urbanicity, and school type (e.g., public/private/charter), as there is substantial evidence that these factors often influence student behaviors. In addition to providing a more holistic picture of research to the readers, this practice of reporting diverse contextual factors could also help with situating the research in prior literature and aid replication or application in a similar context.

Implications for the Fair-ML Community

Despite recent advancements, the field of fair Machine Learning (ML) has been criticized for focusing on algorithmic concerns and mathematical definitions of fairness rather than engaging with the broad set of ethical, social, and political concerns within the contexts in which applications are deployed (Green & Hu, 2018). The interaction between ML systems and social worlds can sometimes lead to effects unanticipated from a purely technical perspective. It is time for application domains like EdTech to actively contribute by bringing our nuanced challenges to the multidisciplinary conversation around fairness (Holstein & Doroudi, 2019). As Selbst et al. (2019) explain, the popular ML approach of abstraction—abstracting out domain-specific aspects of a problem—risks rendering ML ineffective when used to define fairness and develop fair-ML algorithms in a social context. Instead, they argue, fairness requires a nuanced understanding of the social context, its politics, and all the actors involved. In this study, we investigate how the implications of student behavior and motivation within a learning system may be dependent on social and group factors. Specifically, we identify which subpopulations need particular attention by identifying potential blind spots in the generalizability of past results. This work is necessary in order to understand how to make EdTech systems fairer and foster more equitable student outcomes.

A popular strategy to mitigate bias in the industry is to collect more training data (Holstein et al., 2019). However, collecting more rich and complex data in domains like education may not always be feasible. Specifically, collecting fine-grained data on social context and individual demographics can be difficult in education due to student privacy concerns. This study demonstrates a potential workaround to this challenge by collecting coarser-grained school-level demographics data.

Lastly, ML-based sociotechnical systems need to recognize and adapt to the changing social circumstances in the context in which they are deployed—challenging the current practice of treating historical data as the ground truth. For instance, Schofield (1995) reported that students in urban schools skipped lunch and stayed after school to use an intelligent tutoring system—not a common pattern 25 years later. Patterns of interaction and properties of students may change even over a shorter period of time, and may change as a result of using the systems we develop. Moreover, continuing to make predictions using the learning from past data could also impede students' progress in developing interest. To illustrate this, take the example of extrinsic and intrinsic motivation discussed in this paper. The goal of interventions in education is to build student interest in the subject matter, hopefully to the extent that activities started out with extrinsic motivations lead to the student developing an inherent interest in the content. Even when an adaptive EdTech system is trained on data from the same population where it is applied, it may fail to adapt to the improving conditions as a result of the technological intervention if the algorithm stays blind to these changes. For example, continuing to motivate students with extrinsic tangible rewards over increasing student autonomy to attempt challenging problems.

Limitations and Future Work

This paper only considered a small number of sociocultural variables (albeit more than are commonly seen in research on help-seeking or motivation within EdTech). We acknowledge that there are many other sociocultural 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 influence student choices. Similarly, social conditions in school and at home can systemically influence intrinsic and extrinsic motivation. 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 and the value of intrinsic vs. extrinsic motivation.

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) (The Texas Tribune, 2018) 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.

Another potential limitation to our findings is seen in this paper's lack of explicit consideration of gender. 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 still help us understand the role of demographic factors in understanding motivation and help-seeking behaviors.

In addition, it is important to note that the findings were obtained in a single EdTech system, geographical region, and point in time. When using findings of this nature, it is important not just to consider whether the findings are substantial in effect size, but whether they continue to apply. For instance, the impacts of the variables studied here may change between regions (where variables such as limited English proficiency may correspond to different population groups), or over time, as society changes. In general, work investigating the impact of demographics on students' responses to EdTech must be sensitive to the contextual applicability of the phenomena being studied. This indicates that research of the nature presented here must continue to occur over time as well as across learner groups, if we intend our EdTech systems to be effective for all of the learners using them.

The psychological constructs studied in this work—like others that are important to EdTech design—are complex and nuanced. Even though many motivational constructs are studied individually, in practice, they are most likely to co-occur and interact with each other (Renninger et al., 2018). This paper does not represent a complete or comprehensive study on the role of motivation and help-seeking in student performance and identity. However, its finding indicates the importance of future work to consider broader social factors around these constructs when incorporating them in design. In this work, we assume that an outcome is inequitable if particular student groups are observed to be advantaged or disadvantaged by the system usage. We acknowledge that fairness can be conceptualized in other ways too. We hope that our work can contribute to the emerging discussion on fairer EdTech research, design, and development.

Conclusion

In order to close the opportunity gap, we must improve the learning experiences for *all* students who use EdTech systems. This study attempts to answer calls to be socially responsible and accountable in Ed Tech (e.g., Porayska-Pomsta & Rajendra, 2019), which has historically shown considerable social distance between its developers and the students that they want to serve. In order to identify the potential blind spots that lead to inequitable student outcomes, we suggest a method for explicitly identifying the varied needs of student subgroups even when data is unavailable at the student level. We make these recommendations within the context of help-seeking behaviors, which are an important part of self-regulation, and student autonomy more broadly, but which is also a behavior that may be particularly susceptible to cultural differences (i.e., Ogan et al., 2015). Intelligent Tutoring Systems like Reasoning Mind Foundations provide an 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 EdTech systems. Similarly, student autonomy plays an important role in the development of their interest and motivation in learning. In this paper, we demonstrate that school-level demographics can have a significant influence on the relationships between students' help-seeking behavior, their motivations, and student outcomes. In doing so, we question the implicit assumption that complex constructs like these can be considered without also considering student context. This calls for greater consideration within our field of social, cultural, and economic influences on student choices and outcomes (cf. Baker et al., 2019).

Amidst the mixed results from empirical studies on the effectiveness of hints, Aleven and colleagues (Aleven et al., 2016) continue to recommend the use of hints in EdTech systems 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 processes 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. Similarly, we recommend EdTech designers and researchers consider the role of students' demographic contexts while making design choices to motivationally enhance their systems.

This is not to say that there are not both practical and definitional issues in doing so. 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 and of intrinsic and extrinsic motivations. For instance, research should consider the ways in which help-seeking might be interpreted as an imposition or as an admission of failure or how value is attached to autonomy over control for extrinsic motivation, since, as we discussed in Sect. 2, these interpretations likely vary from one culture to another. By considering demographics in our research on these constructs—and on SRL in general—we increase the likelihood that our findings will apply to the full diversity of learners using EdTech and related systems today.

Acknowledgements 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.

Funding This research was supported by Cyberlearning award #1623730.

Data availability Available at http://tiny.cc/RMDemog

Code availability Python code is available at http://tiny.cc/RMDemog

Declarations

Conflicts of interest/Competing interests Not applicable.

References

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.

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, 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).

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).

Amabile, T. M., DeJong, W., & Lepper, M. R. (1976). Effects of externally imposed deadlines on subsequent intrinsic motivation. *Journal of Personality and Social Psychology, 34*, 92–98.

Anderson, J. R. (1993). *Rules of the mind*. Lawrence Erlbaum Associates.

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

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. https://doi.org/10.1007/3-540-45108-0_61

Attewell, P. (2001). Comment: The first and second digital divides. *Sociology of Education, 74*(3), 252–259.

Baker, R. S., Ogan, A. E., Madaio, M., & Walker, E. (2019). Culture in computer-based learning systems: challenges and opportunities. *Computer-Based Learning in Context, 1*(1), 1–13.

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: Springer. https://doi.org/10.1007/978-3-642-21869-9_6

Baltes, M. M. (1997). *The many faces of dependency*. Cambridge University Press.

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

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.

Blanchard, E. G., & Mizoguchi, R. (2008). Designing culturally-aware tutoring systems: towards an upper ontology of culture. *Culturally aware tutoring systems (CATS)*, 23–34.

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

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

Butler, R. (1998). Determinants of help seeking: Relations between perceived reasons for classroom help-avoidance and help-seeking behaviors in an experimental context. *Journal of Educational Psychology, 90*(4), 630.

Butler, R. (2006). An achievement goal perspective on student help seeking and teacher help giving in the classroom: Theory, research, and educational implications. *Help seeking in academic settings: Goals, groups, and contexts*, 15–44.

Chambers, T. V. (2009). The" Receivement Gap": School tracking policies and the fallacy of the" achievement gap". *The Journal of Negro Education*, 417–431.

Chavajay, P., & Rogoff, B. (2002). Schooling and traditional collaborative social organization of problem solving by Mayan mothers and children. *Developmental Psychology, 38*(1), 55.

Childs, D. S. (2017). Effects of math identity and learning opportunities on racial differences in math engagement, advanced course-taking, and STEM Aspiration. Ph.D. Dissertation. Temple University.

Connell, J. P., & Wellborn, J. G. (1990). Competence, autonomy and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.), *The Minnesota symposium on child psychology* (Vol. 22, pp. 43–77). Hillsdale, NJ: Erlbaum.

Crossley, S. A., Karumbaiah, S., Ocumpaugh, J., Labrum, M. J., & Baker, R. S. (2020). Predicting math identity through language and click-stream patterns in a blended learning mathematics program for elementary students. *Journal of Learning Analytics, 7*(1), 19–37.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 627–688.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

Deci, E. L., & Cascio, W. F. (1972, April). Changes in intrinsic motivation as a function of negative feedback and threats. Presented at the meeting of the *Eastern Psychological Association*, Boston.

Doroudi, S., & Brunskill, E. (2019). Fairer but not fair enough on the equitability of knowledge tracing. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 335–339).

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.

Finkelstein, S., Yarzebinski, E., Vaughn, C., Ogan, A., & Cassell, J. (2013). The effects of culturally congruent educational technologies on student achievement. In *International Conference on Artificial Intelligence in Education* (pp. 493–502). Springer, Berlin.

Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap? *The High School Journal, 91*(1), 29–42.

Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal, 33*(4), 111–117.

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.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93*, 3–13.

Green, B. & Hu, L. (2018). The Myth in the methodology: Towards a recontextualization of fairness in machine learning. In *Proceedings of the international conference on machine learning: the debates workshop*.

Greenbaum, P. E., & Greenbaum, S. D. (1983). Cultural differences, nonverbal regulation, and classroom interaction: Sociolinguistic interference in American Indian education. *Peabody Journal of Education, 61*(1), 16–33.

Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology, 52*, 890–898.

Hansen, J. D., & Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science, 350*(6265), 1245–1248.

Harackiewicz, J. (1979). The effects of reward contingency and performance feedback on intrinsic motivation. *Journal of Personality and Social Psychology, 37*, 1352–1363.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70*(2), 151–179.

Hoffmann, L., & Häussler, P. (1998). An intervention project promoting girls' and boys' interest in physics. In L. Hoffmann, A. Krapp, K. A. Renninger, & J. Baumert (Eds.), *Interest and learning* (pp. 301–316). Kiel: IPN.

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.

Holstein, K., & Doroudi, S. (2019). Fairness and equity in learning analytics systems (FairLAK). In *Companion proceedings of the ninth international learning analytics & knowledge conference* (LAK 2019).

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudík, M., Wallach, H. (2019). Improving Fairness in Machine Learning Systems: What do Industry Practitioners Need? In *Proceedings of the ACM CHI conference on human factors in computing systems* (CHI'19). ACM.

Howley, I., Kanda, T., Hayashi, K., & Rosé, C. (2014). Effects of social presence and social role on help-seeking 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. https://doi.org/10.1145/2559636.2559667

Huang, X., Craig, S. D., Xie, J., Graesser, A., & Hu, X. (2016). Intelligent tutoring systems work as a math gap reducer in 6th grade after-school program. *Learning and Individual Differences, 47*, 258–265.

Hudley, A. H. C., & Mallinson, C. (2015). *Understanding English language variation in US schools*. New York: Teachers College Press.

Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2016). Making connections: Replicating and extending the utility value intervention in the classroom. *Journal of Educational Psychology, 109*(3), 387–404. https://doi.org/10.1037/edu0000146

Jackson, G. T., Boonthum, C., & McNamara, D. S. (2009). iSTART-ME: Situating extended learning within a game-based environment. In *Proceedings of the workshop on intelligent educational games at the 14th annual conference on artificial Intelligence in Education* (pp. 59–68).

Karumbaiah, S., Ocumpaugh, J., & Baker, R. S. (2019). The influence of school demographics on the relationship between students' help-seeking behavior and performance and motivational measures. *Educational Data Mining (EDM), 4*, 16.

Karumbaiah, S., Lan, A., Nagpal, S., Baker, R. S., Botelho, A., & Heffernan, N. (2021). Using past data to warm start active machine learning: Does context matter?. In *International learning analytics and knowledge conference* (pp. 151–160).

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.

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

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.

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

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

Koestner, R., Ryan, R. M., Bernieri, F., & Holt, K. (1984). Setting limits on children's behavior: The differential effects of controlling versus informational styles on intrinsic motivation and creativity. *Journal of Personality, 52*, 233–248.

Ladson-Billings, G. (2013). Lack of achievement or loss of opportunity. *Closing the opportunity gap: What America must do to give every child an even chance, 11*.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*(3), 355–365.

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*(2), 184.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic rewards: A test of the '"over justification"' hypothesis. *Journal of Personality and Social Psychology, 28*, 129–137.

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: Springer. https://doi.org/10.1007/978-3-642-39112-5_26

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

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: Springer. https://doi.org/10.1007/978-3-540-70987-9_42

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

Miserandino, M. (1996). Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. *Journal of Educational Psychology, 88*, 203–214.

Nelson-Le Gall, S., & Resnick, L. (1998). Help seeking, achievement motivation, and the social practice of intelligence in school. *Strategic help seeking: Implications for learning and teaching* (pp. 39–60).

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.

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.

Paquette, L., Ocumpaugh, J., Li, Z., Andres, J. M. A. L., & Baker, R. S. (2020). Who's learning? using demographics in EDM research. *Journal of Educational Data Mining, 12*(3), 1–30.

Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International conference on user modeling, adaptation, and personalization* (pp. 255–266). Springer, Berlin.

Porayska-Pomsta, K., & Rajendran, G. (2019). Accountability in human and artificial intelligence decision-making as the basis for diversity and educational inclusion. In *Artificial Intelligence and Inclusive Education* (pp. 39–59). Springer, Singapore.

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.

Reeve, J., & Deci, E. L. (1996). Elements of the competitive situation that affect intrinsic motivation. *Personality and Social Psychology Bulletin, 22*, 24–33.

Reich, J., & Ito, M. (2017). From good intentions to real outcomes: Equity by design in learning technologies. *Digital Media and Learning Research Hub*.

Renninger, K. A., Ren, Y., & Kern, H. M. (2018). Motivation, engagement, and interest:"In the end, it came down to you and how you think of the problem". In *International handbook of the learning sciences* (pp. 116–126). Routledge.

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. https://doi.org/10.1080/10508406.2014.883977

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open, 2*(4), 2332858416673968.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*, 450–461.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*, 749–761.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.

Ryan, R. M., & Grolnick, W. S. (1986). Origins and pawns in the classroom: Self-report and projective assessments of individual differences in children's perceptions. *Journal of Personality and Social Psychology, 50*, 550–558.

Ryan, A. M., Shim, S. S., Lampkins-uThando, S. A., Kiefer, S. M., & Thompson, G. N. (2009). Do gender differences in help avoidance vary by ethnicity? An examination of African American and European American students during early adolescence. *Developmental Psychology, 45*(4), 1152–1163.

Ryan, R. M., Stiller, J., & Lynch, J. H. (1994). Representations of relationships to teachers, parents, and friends as predictors of academic motivation and self-esteem. *Journal of Early Adolescence, 14*, 226–249.

Ryan, R. M., & Stiller, J. (1991). The social contexts of internalization: Parent and teacher influences on autonomy, motivation and learning. In P. R. Pintrich & M. L. Maehr (Eds.), *Advances in motivation and achievement* (Vol. 7, pp. 115–149). JAI Press.

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

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

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). ACM.

Sheldon, K. M., & Kasser, T. (1995). Coherence and congruence: Two aspects of personality integration. *Journal of Personality and Social Psychology, 68*, 531–543.

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.

Skaalvik, E. M., & Skaalvik, S. (2013). School goal structure: Associations with students' perceptions of their teachers as emotionally supportive, academic self-concept, intrinsic motivation, effort, and help seeking behavior. *International Journal of Educational Research, 61*, 5–14.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529.

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. https://doi.org/10.3233/JAI-2011-021

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

Subotzky, S., & Prinsloo, P. (2011). Turning the tide: A socio-critical model and framework for improving student success in open distance learning at the University of South Africa. *Distance Education, 32*(2), 177–193.

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: Springer. https://doi.org/10.1007/978-3-642-39112-5_25

Tessler, R. C., & Schwartz, S. H. (1972). Help seeking, self esteem, and achievement motivation: an attributional analysis. *Journal of Personality and Social Psychology, 21*(3), 318–326.

Texas Education Agency. (2018a). State and school district summary. Retrieved from http://www.texaseducationinfo.org/infopage/Summary_Report_Glossary.pdf. Accessed 26 Feb 2019.

Texas Education Agency. (2018b). District Type Glossary of Terms. Retrieved from https://tea.texas.gov/acctres/analyze/1617/gloss1617.html#Major20Urban. Accessed 26 Feb 2019.

The Texas Tribune. (2018). State and School District Summary. Retrieved from https://www.texastribune.org/2018/08/24/texas-schooldistricts-a-f-grades-takeaways/. Accessed 26 Feb 2019.

Tsai, Y. S., & Gasevic, D. (2017). Learning analytics in higher education---challenges and policies: a review of eight learning analytics policies. In *LAK'17*. ACM.

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

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.
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.
Vallerand, R. J., & Reid, G. (1984). On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. *Journal of Sport Psychology, 6*, 94–102.
VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227–265.
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.
Wilkins, J. L. (2004). Mathematics and science self-concept: An international investigation. *The Journal of Experimental Education, 72*(4), 331–346.
Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology, 70*, 767–779.
Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education, 33*(2/3), 153–169.
Yudelson, M., Fancsali, S., Ritter, S., Berman, S., Nixon, T., & Joshi, A. (2014, July). Better data beats big data. In *Educational Data Mining 2014*.
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.
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.
Zimmerman, B. J. (1985). The development of "intrinsic" motivation: A social learning analysis. *Annals of Child Development*, 117–160. Greenwich, Conn. JAI.

Authors and Affiliations

Shamya Karumbaiah[1] · Jaclyn Ocumpaugh[1] · Ryan S. Baker[1]

Jaclyn Ocumpaugh
jlocumpaugh@gmail.com

Ryan S. Baker
ryanshaunbaker@gmail.com

[1] Penn Center for Learning Analytics, University of Pennsylvania, Philadelphia, PA, USA