Middle School Engagement with Mathematics Software and Later Interest and Self-Efficacy for STEM Careers

Jaclyn Ocumpaugh, Maria O. San Pedro, Huei-yi Lai, Ryan S. Baker, Fred H. Borgen

Abstract:

Research suggests that trajectories towards careers in Science, Technology, Engineering and Mathematics (STEM) emerge early and are influenced by multiple factors. This paper presents a longitudinal study, which uses data from 76 high school students to explore how a student's vocational self-efficacy and interest are related to his or her middle school behavioral and affective engagement. Measures of vocational self-efficacy and interest are drawn from STEM-related scales in CAPAExplore, while measures of middle school performance and engagement in mathematics are drawn from several previously-validated automated indicators extracted from logs of student interaction with ASSISTments, an online learning platform. Results indicate that vocational self-efficacy correlates negatively with confusion, but positively with engaged concentration and carelessness. Interest, which also correlates negatively with confusion, correlates positively with correctness and carelessness. Other disengaged behaviors, such as gaming the system, were not correlated with vocational self-efficacy or interest, despite previous studies indicating that they are associated with future college attendance. We discuss implications for these findings, which have the potential to assist educators or counselors in developing strategies to sustain students' interest in STEM-related careers.

Keywords: STEM, affect, engagement, career self-efficacy, career interest

Middle School Engagement with Mathematics Software and Later Interest and Self-efficacy for STEM Careers

INTRODUCTION

Concerns about an insufficient workforce in science, technology engineering, and mathematics (STEM) fields have spurred a growing interest in the relationships between STEM learning and career trajectories (Cantrell & Ewing-Taylor, 2009; Hayden et al., 2011; Wang, 2013). Research in this area often relies on the Social Cognitive Career Theory (SCCT), which emphasizes the role that self-efficacy and interest play in vocational trajectories. Meanwhile, the emerging literature on behavioral and affective engagement suggests that these constructs are powerful predictors of long-term academic achievement (D'Mello et al., 2008; Liu et al., 2013; Pardos et al., 2014; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010; Pekrun, Goetz, Titz, & Perry, 2002; San Pedro et al., 2013). Such findings suggest that engagement constructs, which were not initially incorporated in SCCT or other core theories of career development, may help to advance our understanding of how vocational trajectories initially develop.

This study brings together these two literatures, examining the extent to which self-efficacy and vocational interest measures, long established as important in career theory models, may be related to measures not typically discussed in that literature, namely measures of behavioral and affective engagement. We do this by comparing high school measures of self-efficacy and interest in STEM careers to measures of affective and behavioral engagement during the same students' earlier interactions with an online tutor for middle-school mathematics.

Established theoretical models such as Social Cognitive Career Theory (SCCT) suggest that secondary students' *self-efficacy, interest*, and *outcome expectations* (Lent et al., 1994) interact with *environmental factors* (i.e., supports and barriers) and *individual attributes* (e.g., personality traits) which subsequently influence career aspirations, decisions (Lent et al., 1994), and behaviors (Betz & Hackett, 1981; Hackett & Betz, 1981).

The SCCT emphasizes the interplay between environmental and individual factors that contribute to the academic and career choices students make (Lent & Brown, 2006), and a considerable body of research has examined how these factors vary across demographics. In general, students' views on careers and work are deeply connected to their relationships and community (Blustein, 2011), meaning they may perceive barriers related to their ethnicity, race, socioeconomic status, family background and gender. For example, career gender-typing has been shown to interact with measures of career-efficacy, influencing whether a student plans for or explores certain fields (Turner & Lapan, 2002), and experiences of classism has been shown to shape goals related to economic survival rather than other career goals (Blustein et al., 2015; Wang 2013). As these factors interact, experiences that impact competency beliefs are thought to influence students' interest formation. The SCCT also emphasizes the importance of learning experiences, which can encourage students to intentionally engage with an activity (e.g., Lent et al, 1994), scaffolding students' self-efficacy and interest (Betz & Hackett, 1981; Hackett & Betz, 1981). As such, early measures of student engagement may help us to identify emerging career trajectories.

Theoretical discussions of student engagement emphasize its multidimensionality (Fredericks et al., 2004; Reschly & Christenson, 2012). Within this paper, we focus on affective and behavioral engagement (e.g., Fredericks et al., 2004), including the affective states known as academic emotions, which appear to influence student learning (Craig et al., 2004; Pardos et al., 2014; Rodrigo et al., 2009). This focus is motivated by research on vocational trajectories that suggests that positive emotions seem to improve self-efficacy by eliciting memories of success, while negative emotions may lower self-efficacy through a similar process (Bandura, 1997), and by research in the learning sciences that suggests that positive and negative emotions differ in terms of the type of learning they promote (see discussion in Pekrun, 2011).

These issues are investigated longitudinally by following 76 learners from middle school to high school to compare how their early engagement is related to their high school self-efficacy and interest. In order to analyze their middle-school engagement, we take advantage of data from online learning technologies, which have become an increasingly prominent part of classroom learning and homework in American schools. These platforms support students in acquiring key concepts and skills through providing automated hints and feedback on student errors (Koedinger & Corbett, 2006) while tracking student performance in close detail (Baker & Yacef, 2009) and providing teachers with detailed reports on what students know and where they need support (Feng & Heffernan, 2005). With the emergence of the educational data mining/learning analytics community (Baker & Yacef, 2009), researchers are now able to use this rich data to infer a wide range of constructs from student interactions within the system. These include constructs typically measured in the classroom (e.g., learning and inquiry skills), but also those which teachers typically observe but rarely attempt to measure. For instance, models have been published which can replicate human judgment of a variety of constructs related to behavioral and affective engagement (see reviews in Baker & Rossi, 2013; Baker & Ocumpaugh, 2014).

In this study, we use models of behavioral and affective engagement that were developed for ASSISTments, a widely-used online educational platform for middle school mathematics. These models have been shown to achieve population validity for diverse students (Ocumpaugh et al., 2014) and have demonstrated relationships with later college attendance (San Pedro et al.,

2013) and choice of college major (San Pedro et al., 2014). This study expands on that work by investigating the relationships between these measures and high school levels of self-efficacy and interest.

Specifically, we apply validated models of *boredom*, *confusion*, *engaged concentration* (the emotion associated with Csikszentmihalyi's (1990) construct of *flow*), *frustration*, *carelessness*, *gaming the system* (intentional misuse of an online learning platform in order to succeed without learning – Baker et al., 2004), and *off-task behaviors* to interaction data produced by these 76 students during their middle-school use of ASSISTments (and recorded in system log files). We then compare the degree to which a student displays these indicators to measures of their vocational interest and self-efficacy obtained from CAPAExplore (Betz & Borgen, 2010) during their junior year of high school. In doing so, this study attempts to contribute to the literature on STEM career development by expanding our understanding of the factors associated with the trajectory of students' vocational interest and self-efficacy.

METHODS

ASSISTments System

This study is conducted in the context of ASSISTments, a free web-based tutoring system for middle school mathematics developed by Worcester Polytechnic Institute (WPI). ASSISTments assesses a student's knowledge while assisting them in learning, providing teachers with detailed reports on the skills each student knows (Razzaq et al., 2005). When students answer correctly, they proceed to the next problem. When they answer incorrectly, the system scaffolds instruction by dividing the problem into component parts, stepping students though each before returning them to the original problem. Once the original problem is correctly answered, the student advances to the next problem. ASSISTments is now used by approximately 50,000 students a year, across the United States but with particularly high usage in Maine and Massachusetts.

Participants

For this study, 284 students who had used ASSISTments as part of their regular middle school instruction in 2007-08 or 2008-09 were contacted during their junior year of high school several years later. Students completed a 12-question survey on course-taking (not analyzed in this paper) during their regular classroom instruction time. Teachers at the high school distributed information about this study to students, who were invited to complete an online survey through the CAPAExplore platform outside of regular school hours. Of the 284 students offered the opportunity to participate, a total of 76 students chose to complete the survey and received a $25 gift card as compensation.

Perhaps counter-intuitively, it appears that our incentive may have contributed to this somewhat low response rate. In Cook et al.'s (2000) meta-analysis of online survey research, they found that potential participants interpret the availability of an incentive as a signal that the survey will require substantial time or effort, deterring them from starting the survey. By contrast, survey length was not correlated strongly to completion rates. That said, while our response rate (27%) is somewhat low, it is not a level uncommon within online survey research (cf. Cook et al.'s meta-analysis), in part because of the relative non-immediacy of taking a survey online. Fortunately, research on survey methods suggests that higher response rates are not necessarily more likely to produce more representative samples (Krosnick, 1999). Nonetheless, this may represent a limitation in our study.

Demographics. Students participating in this study came from a large, urban school district in New England (United States) where they were enrolled in a high school with a graduation rate that is 20% lower than the state average. Information on individual students' demographic variables was not obtained, but according to state-level data on students at this school, 45% were classified as Hispanic, 18% were classified as African-American, and 81% received free or reduced-price lunches.

CAPAExplore (Dependent Variables)

In this paper, we measure *vocational self-efficacy* and *interest* using the CAPAExplore survey instrument (Betz & Borgen, 2009, 2010, 2015), which has been extensively validated for these purposes (Betz & Borgen, 2009, 2010, 2015; Borgen & Betz, 2011, 2015; Larson, 2012; Rottinghaus & Eshelman, 2015). CAPAExplore contains 482 that question participants about interest and efficacy for specific, representative careers and activities related to those careers. The items map to 27 vocational self-efficacy scales and 35 vocational interest scales, as well as to Holland's (1959, 1997) vocational interest scales: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (abbreviated as RIASEC and also referred to as the Big Six interests). CAPAExplore was chosen for several reasons, including the ease of mapping its items to the National Science Foundation's (2013) definitions of STEM and non-STEM career categories, the focus on clearly separable measures of vocational interest and self-efficacy, and the availability of a highly usable online system (http://www.CAPAExplore.com) that was practical for both students and researchers.

For the purposes of this study, 7 complementary scales that map to the National Science Foundation's (2013) definitions of STEM and non-STEM career categories were identified from each set of scales (self-efficacy and interest): (a) *Mechanical (self-efficacy)/Mechanics and Engineering (interest)*, (b) *Information Technology*, (c) *Science*, (d) *Medical Science*, (e) *Mathematics*, (f) *Accounting*, and (g) *Personal Computing*. These scales were aggregated into two measures: the student's vocational interest in STEM, and their self-efficacy for STEM. As part of this aggregation, scores of specific scales were unitized (using a Z-score, also referred to as the normal distribution) within each dimension across participants. By transforming each scale to a Z-score based on the student's difference from the mean across students, and using the standard deviation across students, according to the formula $= \frac{X-Mea}{SD}$, we can give equal weighting to each specific dimension score for each participant, preventing more highly scored dimensions from having a disproportionate influence on the participant's summarized STEM career self-efficacy or interest. Once scores had been transformed into Z-scores, we computed each student's mean score and standard deviation across categories.

Engagement and Performance Indicators (Independent Variables)

The independent variables in this study consist of 7 measures of affective and behavioral indicators of student engagement and 3 measures of mathematics performance/knowledge indicators. These measures make inferences using data drawn from logs – records of student behavior in a learning system automatically collected over the internet by that learning system. These measures infer student boredom, carelessness, confusion, engaged concentration, gaming the system (intentionally misusing educational software to succeed without learning – Baker et al., 2004), and off-task behavior, as well as indicators of student performance, including knowledge models, the percentage of correctly answered questions, and the total number of actions per student over the course of the year.

Boredom, confusion, frustration, and engaged concentration were chosen as the four affective constructs to be investigated. These four constructs play an important role in academic settings, being some of the most common affective states during learning (D'Mello, 2013), with evidence for substantial correlation to learning outcomes, both in the short-term (Craig et al., 2004), over the span of semesters (Rodrigo et al., 2009; Pekrun et al., 2010), and on standardized examinations (Pardos et al., 2013). Their interplay over time is complex, and interact with student knowledge, behavior, and self-regulatory skills (Baker et al., 2010; D'Mello & Graesser, 2012). Previous research has found that the manifestation of boredom, confusion, and engaged concentration as early as middle school are associated with differing probability of eventual college attendance (San Pedro et al., 2013).

Three indicators of behavioral disengagement were chosen: gaming the system, off-task behavior, and carelessness. Gaming the system is seen when a student attempts to succeed in a learning environment by subverting that learning environment rather than attempting to learn the material (Baker et al., 2004). Examples of gaming behavior in online learning include systematic guessing and clicking through hints at high speed until the system gives the answer. There is evidence that gaming the system in middle and high school is associated with worse learning outcomes (Baker et al., 2004; Cocea et al., 2009; Fancsali, 2013), as well as poorer scores of standardized examinations (Pardos et al. 2013), lower probability of college attendance (San Pedro et al., 2013), and lower probability of STEM major enrollment (San Pedro et al., 2014). Off-task behavior has been studied as an indicator of disengagement for decades. Associated with mild but unstable negative associations with learning in traditional classrooms (Karweit & Slavin, 1982; Goodman, 1990), there is relatively limited evidence for off-task behavior being associated with negative outcomes in online learning (e.g. Baker et al., 2004; Cocea et al., 2009;

Fancsali, 2013; Pardos et al., 2013; San Pedro et al., 2013, 2014), but its historical importance in research on behavior disengagement justifies its continued inclusion in research studies. Carelessness has been defined several ways; we adopt the definition from Clements (1982), which focuses on careless errors, situations where the student has the knowledge to answer correctly but answers incorrectly. Recent research has found that after controlling for student knowledge, carelessness in middle school is associated with lower probability of college attendance and lower probability of majoring in STEM (San Pedro et al., 2013, 2014).

The amount of student system usage is measured as a general proxy for conscientiousness and effort, and student correctness and knowledge are measured as indicators of how successful the student is at mathematics, factors that could reasonably be expected to predict self-efficacy according to the SCCT theoretical model reviewed earlier.

Engagement and Performance Indicators (Independent Variables): Model Development

This study measures middle-school student engagement and performance by analyzing the log file data of students' interactions with ASSISTments. It does so by applying *automatic detectors*, or software *models* developed through educational data mining techniques. In this multi-step process, a *ground truth label* is established (e.g., through human observation, as discussed below) and then modeled through a cross-validation process that ensures generalizability to new students (e.g., Baker & Ocumpaugh, 2014). In this section, we discuss both the development of previously published automated detectors of student engagement and other models of performance.

The automatic detectors (models) used in this study infer student engagement from the log files of student interactions with ASSISTments, using measures developed specifically for use on data from that system (Pardos et al., 2014; Ocumpaugh et al., 2014). These models were developed using standard educational data mining prediction modeling methods (e.g. Baker & Yacef, 2009) using methods for measuring affect and behavioral engagement also applied to around a dozen other learning systems (see reviews in Baker & Rossi, 2013; Baker & Ocumpaugh, 2014). They were verified to function accurately for new students in previous research; full details about this process can be found in those publications (Pardos et al., 2014; Ocumpaugh et al., 2014), but we give a brief summary here.

First, human judgments of student engagement were obtained from systematic direct classroom observations using the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP). These judgments were not used as a variable in the study presented in this paper; instead, they were used to generate the automated detector measures which were used in this paper. BROMP is a momentary time-sampling method (e.g. students are briefly observed, one after another) used by over 150 trained coders in 4 countries to assess student engagement.

BROMP was initially designed for developing automated detectors of student engagement, as in this paper, although the observations can also be analyzed directly, without first aligning them to students' log files (Ocumpaugh et al., 2012, 2015). Under this protocol, each student is coded individually; behavioral and affective categories are recorded separately, but simultaneously, so that a student who is off-task and concentrating is treated differently than one who is on-task and concentrating. BROMP coding is conducted holistically; the coder makes a judgment based on a combination of factors, including posture, gaze, verbalizations, facial expression, and the activity surrounding the student, in line with evidence that humans generally recognize emotion holistically in day-to-day settings. When a coder is uncertain about the student's engagement or affect, the observation is dropped from the data set. In order to ensure reliability, BROMP-certified coders undergo rapid but extensive training that involves discussion of affective theory, in-classroom coding practice, and training in reducing observer effects through body positioning and gaze control. All BROMP-certified coders have been verified to agree 60% better than chance with another BROMP-certified coder under real-world classroom conditions. Further details on BROMP are available in the BROMP 1.0 and 2.0 training manuals (Ocumpaugh et al., 2012, 2015).

As discussed above, the students participating in this study were not directly observed by a BROMP-certified coder. Instead, we used automated detectors of student engagement that were developed using *prediction modeling*, a common educational data mining technique (Baker & Yacef, 2009) that allows for the development of cross-validated models that can be applied to new populations of students. In doing so, we used the models from Ocumpaugh et al., (2014), which were developed using BROMP observations of a similar student population.

As explained in Ocumpaugh et al., (2014), this is a multi-step process. (a) First, observations of student engagement are made by a BROMP-certified coder. (b) Next, observations of each student who was observed are synchronized to the log files of that student's interactions with ASSISTments. (Exact synchronization is facilitated by the Human Affect Recording Tool (HART; Ocumpaugh et al., 2015), an Android app that facilitates BROMP data collection). (c) Once the log files of observed students have been labeled with BROMP observations, clips (20-second excerpts) are extracted from the log files for analysis. (d) Features of these clips (i.e., descriptions of the students' interactions with the system during the time of the observation) are then constructed from the log-file data so that (e) different algorithms can produce models, also called automated detectors, that replicate the judgments made by the BROMP-certified human coder. (f) The best-performing models are evaluated using a cross-validation process that verifies that the inferences made by these models will be generalizable to new populations of students. In this study, we use models validated with data from students in the same city and demographics as the participants in this study, and finally (g) we apply these models to the log file history stored by ASSISTments in order to measure the participants' engagement during middle-school mathematics. Full details on the development and evaluation of these models is given in Pardos et al., (2014) and Ocumpaugh et al., (2014).

In addition to the models generated from BROMP field observations, three other measures were used (also described in Pardos et al., 2014): (1) The *number of actions* each student made within ASSISTments was used as a proxy for the student's total activity working with the mathematics tutor. (2) Each student's percentage of correct responses or (*percent correct*) was used as a measure of their success in mathematics. (3) We also used measures of student *knowledge* (a major determinant of self-efficacy – Bandura, 1997), calculated with the Bayesian Knowledge Tracing algorithm (BKT; Corbett & Anderson, 1995), the most-widely used measure of student knowledge in online learning, with accuracy as good or better than several competing methods (Gong, Beck, & Heffernan, 2010; Pardos et al., 2011). This algorithm infers student knowledge from performance during online learning, and is designed to measure knowledge that is changing while it is being measured. Finally, we use San Pedro, Rodrigo & Baker's (2011) model of carelessness. This measure uses a combination of aspects of student performance to infer how often students make errors that are not due to not knowing the relevant mathematical skill. In brief, if a student knew enough to produce the right answer, but answered quickly and incorrectly and non-systematically, it can be inferred that their error was due to carelessness.

Correlational Analyses

The relationships between engagement within ASSISTments and the CAPAExplore measures of vocational interest and self-efficacy were analyzed using Pearson correlations and associated t-tests of the correlations' statistical significance. This set of analyses involves conducting a substantial number of statistical significance tests; as such, we use a post-hoc control to control for Type I (false positive) error. A False Discovery Rate (FDR) correction, Storey's (2002) q, was selected. False Discovery Rate (FDR) corrections attempt to control the proportion of false positives, achieving a level of conservatism that allows 5% of findings to be false, equal to the original conception of statistical significance. A False Discovery Rate correction is selected instead of a Family-Wise Error Rate (FWER) correction, such as the Bonferroni Correction, due to the much higher level of conservatism found for FWER corrections (Perneger, 1998). Storey's q has the benefits that it takes the significance of all tests into account when considering each individual test and that it does not apply a different significance criterion to different tests. Storey's q gives an alternate probability estimate to p, called q. Storey's q values were calculated, using the standard approach of taking the data set's actual range of p values as the lambda range for calculating Storey's q. Under this paradigm, we can consider a result statistically significant if its q value is below 0.05. In this paper, we treat a result as significant if both its original p value and its q value are below 0.05. q values can be lower than p values in cases where so many results are statistically significant as to make it very improbable that the null hypothesis is true, across the set of tests.

RESULTS

Results for Self-Efficacy

Several of the measures of student engagement and knowledge are statistically significantly associated with students' self-efficacy for STEM careers, as shown in Table 1. In fact, 6 of the

10 tests are statistically significant according to a $p=0.05$ cutoff, an overall result with that would be very unlikely if an overall null hypothesis were true and there were genuinely no correlations different than zero ($p=0.0000027$).

[Insert Table 1 here]

As Table 1 shows, confusion was negatively correlated with career self-efficacy ($r=-0.415$, $t(74)= -3.92$, $p<0.001$). Engaged concentration was positively correlated to self-efficacy ($r=0.292$, $t(74)=2.63$, $p=0.011$), as in the original hypotheses. Carelessness and self-efficacy were positively correlated, ($r=0.237$, $t(74)=2.44$, $p=0.017$), contrary to our original hypothesis. We discuss this unexpected finding further in the discussion section below.

By contrast, boredom ($r=-0.176$, $t(74)= -1.54$, $p=0.128$), gaming the system ($r=0.075$, $t(74)=-0.65$, $p=0.52$), and off-task behavior ($r=-0.165$, $t(74)=-1.44$, $p=0.154$) are not significantly associated with STEM career confidence.

A significant positive relationship is seen between STEM career self-efficacy and student knowledge (estimated using Bayesian Knowledge Tracing) as the student learns with ASSISTments ($r=0.291$, $t(74)=2.62$, $p=0.011$). The proportion of correct actions and the number of actions made by the student also showed positive relationships with STEM career self-efficacy, ($r=0.26$, $t(74)=2.32$, $p=0.023$, and $r=0.254$, $t(74)=2.26$, $p=0.027$, respectively).

Results for Career Interest

The relationships between interest and middle school patterns of learning and engagement are less robust than those found for self-efficacy, but they do suggest a relationship between middle school skills and subsequent (high school) interest (see Table 2). 4 of the 10 tests are statistically significant according to a p=0.05 cutoff, an overall result with that would be very unlikely if an overall null hypothesis were true and there were genuinely no correlations different than zero (p=0.00096).

[Insert Table 2 here]

As with career self-efficacy, the strongest correlate to students' interest in STEM careers among our measures is confusion, r=-0.266, t(71)=-2.33, p=0.023. Engaged concentration has the appearance of a positive trend, but this trend is only marginally significant according to the original p value, t(71)=1.89, p=0.06. Frustration, boredom, and gaming the system were not significantly associated with STEM career interest.

Curiously, carelessness is positively correlated to STEM career interest, r=0.246, t(71)=2.14, p=0.036, just as it was positively correlated with STEM career self-efficacy.

As with self-efficacy, student knowledge in middle school is positively correlated to STEM career interest, r=0.236, t(71)=2.05, p=0.045, as was correctness r=0.241, t(71)=2.09, p=0.04.

DISCUSSION/CONCLUSIONS

In this paper, we correlate measures of students' learning and engagement during their middle school use of ASSISTments to CAPAExplore assessments of their interest and self-efficacy during high school. This longitudinal study, which spans several years of students' career development, complements previous research with these detectors, which have already been used to better understand how engagement and learning are related to whether or not a middle school student will eventually go to college (San Pedro et al., 2013) and what major a college-bound student will eventually select (San Pedro et al., 2014). In doing so, this study joins a growing number of research studies that use log files from online learning to facilitate longitudinal research. By linking log data collected several years ago to other measures collected today, we are able to conduct longitudinal research with greater ease than was feasible in previous decades.

Several findings were statistically significant. One interesting finding was that confusion was negatively correlated with both self-efficacy and interest. Students' experiences of confusion have relatively strong negative associations with both career self-efficacy and career interest—stronger than students' overall knowledge or performance within online learning. Although this was not in our initial set of hypotheses, due to the lack of a clear link between these constructs in the published literature, this result is unsurprising. Although confusion is necessary for learning in some cases (D'Mello et al., 2014; Liu et al., 2013), it is also understandable that confused students would not feel confident about their knowledge. It is interesting that this effect persists so strongly over several years, suggesting that the early feeling of confusion either persists over time due to continued struggle with the material, or continues to color students' attitudes towards a domain over an extended period of time. These results suggest that students who are unable to resolve confusion during middle school mathematics lessons are less likely to be interested in pursuing STEM careers, but they are even less likely to be confident in their STEM skills several years later. As such, confusion is an important early indicator of a student who is at-risk of dropping out of the STEM pipeline. It is worth noting that our findings do not provide evidence that confusion in middle school directly *causes* lower self-efficacy or interest in the future. It is possible as well that both the experience of confusion and later self-efficacy and interest stem from generally lower success with mathematics, or from a student's self-identity related to mathematics and science or other personality traits and attributes. Better understanding these relationships will be a valuable area for future work.

Also unexpectedly, we observed a positive correlation seen between carelessness and self-efficacy, and also between carelessness and vocational interest. These correlations might seem counterintuitive. However, previous research has found that students with high confidence in their math abilities (e.g. self-efficacy) tend to become careless (Clements, 1982). However, this finding may be driven by student knowledge as well. If a student has not shown mastery of that skill, then their mistakes are typically not labeled as careless by the detector. As such, student knowledge may mediate the relationship between carelessness and self-efficacy. It is possible that the relationship between carelessness and interest may not be causal; instead, it can be hypothesized to be due to the connections between carelessness and student knowledge (more knowledgeable students get careless; Clements, 1982; San Pedro et al., 2013) and the connections between student knowledge and career interests. Alternatively, as with any single finding, it is possible that these results are spurious, although this is not the first unexpected finding associated with carelessness (e.g. Clements, 1982; Pardos et al., 2013; San Pedro et al., 2011, 2013, 2014), suggesting that our theoretical understanding of this potentially important construct is still quite limited.

More in line with our original hypotheses, engaged concentration was positively correlated with self-efficacy. The links between self-efficacy and the flow state (Csikszentmihalyi & Schneider, 1989), as well as between engaged concentration itself and future outcomes, combine to make this result plausible. However, the relatively stronger link between engaged concentration and career self-efficacy than between engaged concentration and interest is interesting. These findings may suggest that that flow emerges in a context of achievement, which in turn leads to career self-efficacy. Alternatively, engaged concentration may encourage students to persevere even when material becomes difficult, leading to greater career self-efficacy. Finally, it may be that there is a direct link; students are more confident with material they find engaging. Whatever the explanation, the findings of this paper regarding engaged concentration merit further analysis and investigation.

Interestingly, boredom, frustration, gaming the system, and off-task behavior are not significantly associated with STEM career confidence or vocational interest in STEM, despite previous research that has shown the importance of boredom and gaming the system for other educational outcomes, such as learning (Pardos et al., 2013; Rodrigo et al., 2009), college attendance (San Pedro et al., 2013), and enrollment in STEM majors during college (San Pedro et al., 2014). One explanation for these weak relationships could be that students in our sample may have been more prone to these forms of disengagement if they believed that the practice provided by ASSISTments was unnecessary. Alternatively, in the case of self-efficacy, it may be that a student's poor performance does not reduce their self-efficacy, if the student believes their poor performance is due to a lack of trying (Thomas & Gadbois, 2007; Urdan, 2004; Urdan & Midgley, 2001). Still, these are somewhat surprising findings since frustration in particular might be expected to correlate to self-efficacy in the same way that confusion does. However previous research has suggested that successful students may become frustrated when minor errors trigger ASSISTments to provide them with extra practice (San Pedro et al., 2013). Given these findings, future research should consider whether frustration's relationship to the CAPA measures may be complicated by differences between high and low-achieving students. The lack of connection between boredom and interest is also surprising. One might expect that students who are bored during math class to be less interested in careers that involve mathematics. In fact, this result accords with another null effect: boredom is also not correlated with whether students enroll in STEM majors (San Pedro et al., 2014). One potential explanation is that the boredom some students experience with ASSISTments may not reflect a lack of disinterest with the subject of mathematics, so much as it reflects boredom with schooling or with the ASSISTments system. Minor differences in system design can have large impacts on boredom for some students (Doddannara et al., 2013). It is also possible that boredom with mathematics might not detract from interest in fields that require its use; that is, students might be interested in careers in medicine, for example, even if they don't find mathematics to be intrinsically interesting by itself.

The lack of connection between boredom, gaming the system, and frustration and later indicators of career interest and career self-efficacy does not mean, however, that these factors should not be addressed during middle school since they still impact both learning outcomes and the probability of college enrollment (Pardos et al., 2013; San Pedro et al., 2013). As such, they influence student career outcomes, but, at least in this sample, that influence does not extend to career interest and self-efficacy.

A significant positive relationship is seen between STEM career self-efficacy and student knowledge (estimated using Bayesian Knowledge Tracing) as the student learns with ASSISTments. This finding, that more knowledgeable students in middle school have higher self-efficacy, is perfectly in keeping with theory on self-efficacy. The proportion of correct actions and the number of actions made by the student also showed positive relationships with STEM career self-efficacy. Again, getting mathematics problems correct should lead to higher self-efficacy, as well as completing more problems. Additionally, student knowledge in middle school and correctness are positively correlated to STEM career interest. These results suggest that students who are struggling to learn middle school mathematics (low knowledge, low correctness, high confusion) are less likely to be interested in STEM careers when they get to high school, which is an important finding for designing interventions, since it suggests that early (middle-school) indicators of ability are closely tied to interest levels in high school—a time when students are poised to make post-secondary decisions.

In thinking about these findings, it is important to consider the limitations in the study which produced them. One limitation to the study presented here is the relatively small sample size, of 76 students. Some findings which were not statistically significant may have emerged as statistically significant for a larger sample. Correspondingly, while findings that were statistically significant can be treated as reliable, the estimates of correlation are less precise than would be the case with a larger sample. Beyond the small sample size, this study's focus on urban students in one city in the Northeastern United States presents a clear limitation to how broadly we can generalize our findings. Future work should endeavor to reach a larger and more diverse sample of students. The study's correlational results were modest, in the small to medium range. While this magnitude of correlation is common in the social sciences, it nonetheless warrants caution about over-stating the degree of impact these findings may have for practice; other factors beyond simply middle school engagement contribute to students' self-efficacy and interest several years later.

Nonetheless, the results here suggest that indicators of engagement and successful performance during middle school are correlated to both interest and for self-efficacy in STEM subjects several years later. Beyond expanding theoretical understanding, one possible use of these findings is to give educators and career counselors new, early, high-quality information on students' career trajectories. There have been recent calls for guidance counselors to integrate a range of types of information into their practice with students (Rottinghaus & Eshelman, 2015); this type of data can be a valuable tool for gaining richer understanding than current instruments afford. Guidance counselors already have data on grades and standardized examinations, but grades are lagging indicators (and incorporate factors other than learning, such as behaving according to classroom expectations – Wentzel, 1993), and standardized examinations are also lagging indicators. By contrast, learning and performance data from systems like ASSISTments can be available moments after the student completes a homework set, providing timely data on how students are faring. Career interest and self-efficacy instruments like CAPAExplore are valuable, but time-consuming to administer; as such, they cannot be administered several times a year. By contrast, data on engagement and emotion from a system like ASSISTments is available continually through the year, and across years.

As such, data and models from systems like ASSISTments provides a tool for identifying at-risk students and identifying why a given student is at-risk, giving career counselors and teachers the information that is necessary to select students for interventions. For example, a student who is experiencing considerable confusion may benefit from additional educational support or from interventions that can communicate the value of persisting in the face of confusion (D'Mello & Graesser, 2012; Lee et al., 2011). ASSISTments already provides extensive reports to instructors (Feng & Heffernan, 2005), and work is underway to engineer reports for guidance counselors. Adding information to reports on the potential implications for career self-efficacy and interest could be a valuable additional dimension to these reports. If ASSISTments reports could be used to trigger support mechanisms during middle school, we might improve the chances that more students would be inclined to pursue STEM careers. Ultimately, better understanding the processes that lead to student career interest and self-efficacy will help us to prepare students for the careers of the 21st century and for careers that will lead them to success and fulfillment in their lives.

ACKNOWLEDGMENTS
We would like to thank XXXXXX and the National Science Foundation (award # XXXXX).

REFERENCES

Baker, R. S. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In *Proceedings of the ACM SIGCHI Conference on Human factors in Computing Systems,*1059-1068.

Baker, R. S., Corbett, A. T., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *Proceedings of the International Conference on Intelligent Tutoring Systems,* 406-415.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. In *Proceedings of the International Conference on Intelligent Tutoring Systems.* 531-540.

Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., & Graesser, A.C. (2010) Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. *International Journal of Human-Computer Studies*, 68 (4), 223-241

Baker, R.S.J.d., & Ocumpaugh, J. (2014) Interaction-Based Affect Detection in Educational Software. In R.A. Calvo, S.K. D'Mello, J. Gratch, A. Kappas (Eds.), *The Oxford Handbook of Affective Computing*. Oxford, UK: Oxford University Press.

Baker, R.S.J.d., & Rossi, L.M. (2013) Assessing the Disengaged Behavior of Learners. In Sottilare, R., Graesser, A., Hu, X., & Holden, H. (Eds.) *Design Recommendations for Intelligent Tutoring Systems -- Volume 1 -- Learner Modeling*. U.S. Army Research Lab, Orlando, FL, pp. 155-166, 2013

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, *19*(2), 185-224.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3-17.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84* (2), 191.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28* (2), 117-148.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52(1), 1-26.

Bandura, A., Caprara, G. V., Barbaranelli, C., Gerbino, M., & Pastorelli, C. (2003). Role of Affective Self-Regulatory Efficacy in Diverse Spheres of Psychosocial Functioning. *Child Development*, 74(3), 769-782.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72(1), 187-206.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *British Journal of Educational Psychology*, *81*(1), 78-96.

Betz, N.E., & Borgen, F.H. (2006). *Manual for the Career Confidence Inventory.* Ames, IA: CAPA.

Betz, N.E., & Borgen, F.H. (2009). Comparative effectiveness of CAPA and FOCUS online career assessment systems with undecided college students. *Journal of Career Assessment*, *17*(4), 351-366.

Betz, N. E., & Borgen, F. H. (2010). The CAPA integrative online system for college major exploration. *Journal of Career Assessment, 18,* 317-327.

Betz, N. E., & Borgen (2015). *CAPA Confidence Inventory manual: Foundations, research, and interpretation*. Ames, IA: CAPA.

Betz, N.E., Borgen, F.H., & Harmon, L.W. (2006). Vocational confidence and personality in the prediction of occupational group membership. *Journal of Career Assessment, 14*, 36-55.

Betz, N. E., Borgen, F. H., & Harmon, L. E. (2012). *Skills Confidence Inventory manual: Research, development, and strategies for interpretation*. Mountain View, CA: CPP.

Betz, N.E., & Hackett, G. (2006). Career self-efficacy theory: Back to the future. *Journal of Career Assessment, 14*, 3-11.

Betz, N.E., Harmon, L.W., & Borgen, F. (1996). The relationships of self-efficacy for the Holland themes to gender, occupational group membership, and vocational interests. *Journal of Counseling Psychology*, *43*(1), 90.

Betz, N.E., & Rottinghaus, P. (2006). Current research on parallel measures of interests and confidence for basic domains of vocational activity. *Journal of Career Assessment, 14* (1), 56-76.

Blustein, D. L. (2011). A relational theory of working. *Journal of Vocational Behavior*, *79*(1), 1-17.

Blustein, D.L., Barnett, M., Mark, S., Depot, M., Lovering, M., Lee, Y., Hu, Q., Kim, J., Backus, F., Dillon-Lieberman, K., & DeBay, D. (2012). Examining Urban Students' Constructions of a STEM/Career Development Intervention Over Time. *Journal of Career Development, 40* (1), 40-67.

Blustein, D.L., Kozan, S., Connors-Kellgren, A., & Rand, B. (2015) Social Class and Career Intervention. In P.J. Hartung, M.L. Savickas, and W.B. Walsh (Eds.) *APA Handbook of Career Intervention: Vol. 1, Foundations.* Washington, DC: American Psychological Association.

Borgen, F. H, & Betz, N. (2008). Career self-efficacy and personality: Linking the Career Confidence Inventory and the Healthy Personality Inventory. *Journal of Career Assessment*, *16*, 22–43.

Borgen, F. H., & Betz, N. E. (2008). *Manual for the CAPA Interest Inventory*. Ames, IA: CAPA.

Borgen, F. H., & Betz, N. E. (2011). Integrating self through personality, interests, and self-efficacy. In P. J. Hartung & L. M. Subich (eds.) *Developing Self in Work and Career: Concepts, Cases, and Contexts* (pp. 141-160). Washington, DC: American Psychological Association.

Clements. K. (1982). Careless errors made by sixth-grade children on written mathematical tasks. *Journal for Research in Mathematics Education*, *13*, 136-144.

Cantrell, P., & Ewing-Taylor, J. (2009). Exploring STEM career options through collaborative high school seminars. *Journal of Engineering Education*, *98*(3), 295-303.

Cocea, M., Hershkovitz, A., Baker, R.S.J.d. (2009) The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate? *Proceedings of the 14th International Conference on Artificial Intelligence in Education*, 507-514.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement*, *60*(6), 821-836.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157.

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

Craig, S., Graesser, A., Sullins, J., Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, *29* (3), 241-250.

Csikszentmihalyi, M. (1990). Flow and the psychology of discovery and invention. *New York, NY: Harper Collins.*

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*, *105*(4), 1082

D'Mello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User modeling and user-adapted interaction*, 18(1-2), 45-80.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153-170.

Doddannara, L., Gowda, S., Baker, R.S.J.d., Gowda, S., & de Carvalho, A.M.J.B (2013) Exploring the relationships between design, students' affective states, and disengaged behaviors within an ITS. *Proceedings of the 16th International Conference on Artificial Intelligence & Education*, 31-40.

Dweck, C. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*(1), 36-48.

Fancsali, S. E. (2013). Data-Driven Causal Modeling of" Gaming the System" and Off-Task Behavior in Cognitive Tutor Algebra. In *Proceedings of the NIPS Workshop on Data Driven Education.* Las Vegas, NV, USA. December 10, 2013.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297-327.

Feng, M., & Heffernan, N. T. (2005). Informing teachers live about student learning: Reporting in the ASSISTments system. *Technology, Instruction, Cognition and Learning, 3,* 1–14.

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, *19*(3), 243-266.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. (2004). School engagement: Potential of the concept: State of the evidence. *Review of Educational Research*, *74*, 59-119.

Goodman, L. (1990). *Time and learning in the special education classroom*. Albany, NY: SUNY Press.

Harmon, L. E., Hansen, J. C., Borgen, F. H., & Hammer, A. L. (1994). *Strong Interest Inventory: Applications and technical guide*. Stanford, CA: Stanford University Press.

Hayden, K., Ouyang, Y, Scinski, L., Olszewski, B., & Bielefeldt, T. 2011. Increasing student interest and attitudes in STEM: Professional development and activities to engage and inspire learners. *Contemporary Issues in Technology and Teacher Education*, *11*(1), 47-69.

Hochberg, Y., & Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine*, *9*(7), 811-818.

Holland, J.L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, *6*, 35-45.

Holland, J.L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed). Odessa, FL: Psychological Assessment Resources.

Karweit, N., & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology*, *74*(6), 844.

Koedinger, K. R., & Corbett, A. (2006). Cognitive tutors. *The Cambridge handbook of the learning sciences*, 61-77. Cambridge, UK: Cambridge University Press.

Krosnick, J. (1999). Survey research. *Annual Review of Psychology*, *50*, 537-567.

Larson, L.M., & Borgen, F.H. (2006). Do personality traits contribute to vocational self-efficacy? *Journal of Career Assessment*, *14*(3), 295-311.

Larson, L. M. (2012). Worklife across the lifespan. In E. M. Altmaier & J. C. Hansen (eds.). *The Oxford Handbook of Counseling Psychology* (pp. 128-178). New York: Oxford University Press.

Lee, D.M.C., Rodrigo, M.M.T., d Baker, R.S., Sugay, J.O., & Coronel, A. (2011). Exploring the relationship between novice programmer confusion and achievement. *Affective Computing and Intelligent Interaction,* 175-184.

Lent, R.W., Brown, S.D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*(1), 79-122.

Lent, R. W., Brown, S. D., & Hackett, G. (2002). Social cognitive career theory. *Career Choice and Development*, *4*, 255-311.

Lent, R. W., & Brown, S. D. (2006). On conceptualizing and assessing social cognitive constructs in career research: A measurement guide. *Journal of Career Assessment*, 14(1), 12-35.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30.

National Science Foundation, (2013). Science and Engineering Degrees: 1966–2010. *Detailed Statistical Tables NSF* 13-327. Arlington, VA: National Science Foundation.

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.

Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 117-124.

Pekrun, R. (2011) Emotions as Drivers of Learning and Cognitive Development. In R.A. Calvo, S.K. D'Mello (Eds.) *New Perspectives on Affect and Learning Technologies.* pp. 23-39. New York, NY: Springer.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91-105.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. British Medical Journal, *316* (7139), 1236-1238.

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In Christenson, S.L., Reschly, A.L., & Wylie, C. (Eds.) *Handbook of research on student engagement* (pp. 3-19). New York, NY: Springer.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261.

Rodrigo, M. M. T., Baker, R. S., Jadud, M. C., Amarra, A. C. M., Dy, T., Espejo-Lahoz, M. B. V. Lim, S.A.L., Pascua, S.A.M.S., Sugay, J.O., & Tabanao, E. S. (2009). Affective and behavioral predictors of novice programmer achievement. *ACM SIGCSE Bulletin, 41*(3), 156-160.

Rottinghaus, P. J., Larson, L. M., & Borgen, F. H. (2003). The relation of self-efficacy and interests: A meta-analysis of 60 samples. *Journal of Vocational Behavior*, *62*(2), 221-236.

Rottinghaus, P. J., Jenkins, N., & Jantzer, A. M. (2009). Relation of depression and affectivity to career decision status and self-efficacy in college students. *Journal of Career Assessment*, *17*(3), 271-285.

Rottinghaus, P.J., & Eshelman, A.J. (2015). Integrative approaches to career intervention. In P.J. Hartung, M.L. Savickas, & W.B. Walsh (Eds.) *APA handbook of career intervention, Vol 2: Applications.* Washington, DC: American Psychological Association. 25-39.

Rottinghaus, P.J., & Van Esbroeck, R. (2011). Improving person-environment fit and self-knowledge. In P.J. Hartung, & L.M. Subich (Eds.) *Developing self in work and career: Concepts, cases, and contexts.* Washington, DC: American Psychological Association, 35-52.

San Pedro, M. O. C. Z., Baker, R. S., & Rodrigo, M. M. T. (2011). Detecting carelessness through contextual estimation of slip probabilities among students using an intelligent tutor for mathematics. *Proceedings of the International Conference on Artificial Intelligence in Education,* 304-311.

San Pedro, M.O.C.Z., Rodrigo, M.M.T., & Baker, R.S. (2011). The relationship between carelessness and affect in a Cognitive Tutor. *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, 306-315.

San Pedro, M.O.Z., Baker, R.S., Bowers, A.J., & Heffernan, N.T. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. *Proceedings of the 6th International Conference on Educational Data Mining*, 177-184.

San Pedro, M.O.Z., Ocumpaugh, J. L., Baker, R.S., & Heffernan, N.T. (2014). Predicting STEM and Non-STEM College Major Enrollment from Middle School Interaction with Mathematics Educational Software. *Proceedings of the 7th International Conference on Educational Data Mining*, 276-279.

Schaub, M. (2004). Social cognitive career theory: Examining the mediating role of sociocognitive variables in the relation of personality to vocational interests. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 64* (7-A), 2463.

Schaub, M., & Tokar, D. M. (2005). The role of personality and learning experiences in social cognitive career theory. *Journal of Vocational Behavior*, *66*(2), 304-325.

Storey, J.D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, *64*, 479-498.

Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy*, *24*(1-2), 23-50.

Thomas, C.R., & Gadbois, S.A. (2007). Academic self-handicapping: The role of self-concept clarity & students' learning strategies. *British Journal of Educational Psychology*, *77*(1), 101-19.

Tobin, T. J., & Sugai, G. M. (1999). Using sixth-grade school records to predict school violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders*, *7* (1), 40-53.

Todd, A. W., Horner, R. H., & Sugai, G. (1999). Self-Monitoring and Self-Recruited Praise Effects on Problem Behavior, Academic Engagement, and Work Completion in a Typical Classroom. *Journal of Positive Behavior Interventions*, *1*(2), 66-122.

Urdan, T. (2004). Predictors of Academic Self-Handicapping and Achievement: Examining Achievement Goals, Classroom Goal Structures, and Culture. *Journal of Educational Psychology, 96*(2), 251.

Urdan, T., & Midgley, C. (2001). Academic self-handicapping: What we know, what more there is to learn. *Educational Psychology Review*, *13*(2), 115-138.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, *21*(3), 209-249.

Wang, X. (2013). Why Students Choose STEM Majors Motivation, High School Learning, and Postsecondary Context of Support. *American Educational Research Journal*, *50*(5), 1081-1121.

Wentzel, K.R. (1993) Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85* (2), 357-364.

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*, *45*(9), 1036-1058.

Table 1. Correlations between engagement in ASSISTments and STEM career self-efficacy in high school. * denotes that a result is significant when taken individually and after Storey's Q correction.

Measures within ASSISTments		Pearson Correlation	p-value	q-value
Engagement	**Confusion**	-0.415*	<0.001	<0.001
	Boredom	-0.176	0.128	0.028
	Carelessness	0.273*	0.017	0.007
	Engaged Concentration	0.292*	0.011	0.006
	Gaming the System	0.075	0.52	0.089
	Frustration	-0.193	0.095	0.023
	Off-Task	-0.165	0.154	0.029
Learning	**Correctness**	0.26*	0.023	0.008
	Number of Actions	0.254*	0.027	0.008
	Student Knowledge	0.291*	0.011	0.006

Table 2. Correlations between Engagement in ASSISTments and STEM career interest in high school. * denotes that a result is significant when taken individually and after Storey's Q correction.

STEM Career Interest	Pearson Correlation	p-value	q-value
Confusion	-0.266*	0.023	0.003
Boredom	-0.095	0.426	0.013
Carelessness	0.246*	0.036	0.003
Engaged Concentration	0.219	0.062	0.004
Frustration	-0.141	0.235	0.036
Gaming the System	0.091	0.442	0.013
Off-Task	-0.182	0.124	0.005
Correctness	0.241*	0.04	0.003
Number of Actions	0.210	0.075	0.004
Student Knowledge	0.236*	0.045	0.003