Middle School Learning, Academic Emotions and Engagement as Precursors to College Attendance

Maria Ofelia Clarissa Z. San Pedro

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

ABSTRACT

Middle School Learning, Academic Emotions and Engagement as Precursors to College Attendance

Maria Ofelia Clarissa Z. San Pedro

This dissertation research focuses on assessing student behavior, academic emotions, and knowledge within a middle school online learning environment, and analyzing potential effects on students' interests and choices related to decisions about going to college. Using students' longitudinal data ranging from their middle school, to high school, to postsecondary years, this dissertation uses quantitative methodologies to investigate antecedents to college attendance that occur as early as middle school. The dissertation asks whether student behavior, academic emotions, and learning as early as middle school can be predictive of college attendance years later. This is investigated by developing predictive and structural models of said outcomes, using assessments of learning, emotions and engagement from student interaction data from an online learning environment they used in their middle school curriculum. The same middle school factors are also assessed with self-report measures of course choices, interests in college majors and careers formed when they were in high school. The dissertation then evaluates how student choices and interests in high school can mediate between the educational experiences students have during middle school and their eventual college attendance, to give a fuller illustration of the cognitive and non-cognitive mechanisms that students may experience throughout varied periods in school. Such understanding may provide educators with actionable information about a students' in-depth experiences and trajectories within the college pipeline.

Table of Contents

List of Tables

List of Figures

ACKNOWLEDGMENTS

As the adage usually goes, *"No one can do it alone."* And nowhere has this been truer than in working towards a PhD degree. I owe a huge debt of gratitude to individuals who have supported and inspired me during the course of my PhD journey, and without whom, this dissertation would not have come to its fruition. I appreciate you all.

To Dr. Ryan Shaun Baker, my PhD sponsor and advisor, life adviser and career mentor, I am deeply honored to be one of your first PhD advisees. Thank you for the inspiration and motivation you have shown throughout my studies and research work, for the unyielding support and guidance as I worked on this dissertation. Thank you for taking a chance on me – in cultivating my interest and desire to learn and grow as a researcher in this interdisciplinary field of learning sciences, data science and education – areas which I intend to embark on going forward. Thank you for the knowledge and insights you have imparted regarding learning analytics and educational research, and the numerous research opportunities you have provided in my training as a researcher in the field.

To members of my dissertation committee who were extremely helpful and very generous with their time and expertise in improving this dissertation research – thank you for imparting valuable feedback and lessons that I will keep as I learn and grow further as an educational researcher. I would like to thank Dr. James Corter for his support in advising me on my methods and data analyses. Together with Dr. Corter, I would also like to thank Dr. Elizabeth Tipton who has helped shape this dissertation from the time I proposed this study. I would also like to thank Dr. Alex Bowers and Dr. Luc Paquette, who were gracious with their time, shared their expertise, and provided further feedback that have made me think more deeply about this research and its future directions.

In addition, I would also like to thank Dr. Neil Heffernan and the ASSISTments team back in Worcester Polytechnic Institute where I started my PhD studies, and where my involvement with ASSISTments grant projects has led me to formulate the research questions for this study. Thank you for your generous support.

To the members of the Educational Data Mining Lab in Teachers College – Dr. Jaclyn Ocumpaugh, Shimin Kai, Mia Almeda, Elle Wang, Stefan Slater, Miggy Andres, Yang Jiang, Jeanine DeFalco – thank you for all the encouragement, camaraderie, and just being there during moments of distress and exultation over the past few years living in New York City and doing research at Teachers College. I treasure them all.

And lastly, to my family and friends, both here in the US and back home in the Philippines, your unconditional love and support have sustained me. Thank you.

DEDICATION

For Mom and Dad.

CHAPTER I.

INTRODUCTION

Background

College attendance and completion are key steps towards career success for many learners. Getting a college degree is related to a higher chance of getting a job (Carnevale, Smith, & Strohl, 2010) and higher levels of social and economic attainment (Reardon, Baker, & Klasik, 2012). This begins with students aspiring to attend or enroll in college. Students go through a longitudinal and complex process of developing these aspirations over the course of elementary, middle and high school (Cabrera & La Nasa, 2000; Hossler, Braxton, & Coopersmith, 1989). Learning opportunities in school and effective guidance and support from educators help shape these student aspirations that influence students' plans to go to college, their academic preparation for college, and their eventual choice to enroll in college.

However, along this pathway to college, students have varied educational experiences that result to either fully realizing this pathway or falling off this pathway. College access among diverse student groups remains inequitable. Minority, low income or first-generation students are usually underrepresented among full-time four-year college students, compared to White, middle- or high-income students (Cabrera & La Nasa, 2001; Pathways to College Network, 2004). Many factors contribute to the differences in college access and attendance among these student groups, one of the most important factors being academic preparation. Many high school graduates may fail to successfully transition to college – dropping out in their first year of college or not enrolling in college at all. Many students find themselves either unprepared and lacking the skills needed to enter college, or they think about going to a selective college but fail to get appropriate support in planning how to achieve this (Balfanz, 2009; Camblin, 2003).

Indicators may begin to manifest as early as middle school, in terms of greater academic failure such as failing grades (National Middle School Association, 2002; Neild, 2009), in terms of decreasing motivation (Anderman & Maehr, 1994), or in terms of extreme forms of disengaged behavior (low attendance, tardiness and misconduct) which result in disciplinary referrals (Tobin & Sugai, 1999). Such changes can eventually translate to academic decisions in the long-run, such as going or not going to college.

Statement of the Problem

Cabrera, La Nasa, and Burkum (2001) showed that the strongest predictors of college access are parental involvement, expectations, and support; academic achievement; financial aid; socioeconomic status (SES); participation in college preparatory classes; academic aspirations; peer and school expectations; and access to guidance counseling. This list of predictors includes student characteristics but also strategies that contribute to students' access to college and success in college. According to Social Cognitive Career Theory (SCCT, Lent, Brown, & Hackett, 1994), academic and career choices are shaped throughout middle school and high school by environment supports and barriers, where higher levels of interest emerge within contexts in which the individual has higher self-efficacy and outcome expectations, and these interests lead to the development of intentions or goals for further exposure and engagement with the activity (Lent, Brown, & Hackett, 1994). And while strong predictors of college access that involve the students' background and their environment, such as socioeconomic status (SES) and parental involvement, can form the supports or barriers for their academic and career choices (e.g. college attendance) mentioned in SCCT, they do not fully explain these choices. The processes at the core of SCCT (i.e. self-efficacy, outcome expectations, interest, choice) suggest opportunities for interventions that can be challenging if the factors to consider are not actionable. Hence, it would be valuable for a student's academic preparation to include a pathway or guidance towards academic and career success that considers actionable factors influencing their self-efficacy, interests and choices.

One factor that can influence students' academic preparation and eventual choice for college is their behavior in school. As mentioned, disengaged behavior in students can manifest in the form of low attendance, tardiness or misconduct. But these behaviors are fairly strong displays of disengagement. By the time these indicators are commonplace, students may be in such a precarious situation that many interventions may fail. Many studies also show that family background, financial resources, and prior family academic achievement have strong, significant impacts on where students find themselves after high school. Similarly, however, these factors are not actionable in terms of being directly changeable by school-based interventions. In general, current models about successful access to postsecondary education may be insufficient to help educators identify which students are on track and which need further support (Lent, et al., 2008).

For that reason, this dissertation attempts to answer Bowers' (2010) call to identify early, less acute signals of disengagement, the sort that occur when students' engagement is still malleable – i.e., amenable to intervention. Specifically, this study investigates antecedents to college attendance that occur during middle school, using assessments of engagement and disengagement to better understand how these factors interact, so that possible paths to re-engagement can be developed before students develop more serious academic problems. Evaluating these factors as early as middle school may provide educators with information about a student's educational progress and facilitate guiding that student towards academic success.

Objectives of the Study

This dissertation research aims to assess cognitive and non-cognitive factors based on interactions with an online learning environment in middle school classrooms, and to analyze their potential effects on students' interests and choices that eventually influence their decisions to go to college. It aims to show that cognitive and non-cognitive factors such as knowledge, engagement and academic emotions in middle school play an essential early role in the processes described in SCCT. In SCCT, students' initial vocational interests are modified by their self-efficacy, attitudes, and goals towards career development (i.e. college enrollment, career interest). Self-efficacy, attitudes and goals are themselves influenced by the student's learning and engagement when encountering the increasingly challenging content in middle school (Baker et al., 2008; McQuiggan, Mott, & Lester, 2008) – for example, poor learning reduces self-efficacy whereas successful learning increases self-efficacy (cf. Bandura, 1997). Students' engaged/disengaged behaviors and academic emotions (emotions that students experience during learning and classroom instruction) are common in classrooms and have been found to influence learning outcomes (McQuiggan, Mott, & Lester, 2008; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). As such, a student's engaged or disengaged behavior, academic emotions, and learning during middle school may be indicative of their developing interest in career domains which may in turn influence their choice to attend college.

In recent years, educational technologies have been used by researchers in exploring educational constructs related to student learning. Educational technologies or software have been a valuable instrument in conducting educational research, providing educational tasks and content to their users (e.g. students, teachers, administrators) that enable researchers to examine different phenomena within these environments. One resource in investigating a user's educational experience is the interaction data that can be acquired from educational software. Through these systems, students produce a series of actions as they complete the learning activities, creating a rich source of data that can assess whether a student's choices and behaviors translate into learning, complementing traditional performance assessments such as standardized tests (Clarke-Midura & Dede, 2010). This dissertation leverages these resources in addressing its research questions about exploring actionable, fine-grained measures of engagement and performance that start as early as middle school, and investigating whether these factors predict long-term student outcomes several years after using a learning system.

Using students' longitudinal data ranging from their middle school, to high school, to postsecondary years, quantitative methodologies were used in this dissertation to investigate malleable antecedents to college attendance that occur as early as middle school. By malleable antecedents, these pertain to factors that can be changed by interventions from educators. These malleable factors may include student behaviors or skills, teacher practices, curricula, school programs or policies. In this study, the malleable factors consisted primarily of students' academic emotions, behavior, and knowledge during middle school computer-based math learning. These factors encompass the cognitive, behavioral and emotional dimensions of student engagement (Fredricks, Blumenfeld, & Paris, 2004) that can be shaped by interventions addressing negative emotions (ex. boredom, anxiety, etc.) and disengaged behaviors to improve learning and achievement.

The dissertation aims to develop three models that examine (1) how students' eventual attendance in college as well as selectivity of college attended can be associated with malleable factors within the context of computer-based math learning as early as middle school, (2) how choice of a college major when students enroll in college can also be associated with malleable factors within middle school computer-based math learning, and (3) how student domain interests and course choices in high school can mediate between the educational experiences students have during middle school computer-based math learning and their eventual college attendance choices. This dissertation leverages existing data acquired from traditional research methods as well as methodologies from machine learning and student modeling to assess the constructs of interest used in the outcome models.

The dissertation research investigates malleable antecedents to college attendance outcomes such college enrollment, selectivity of college attended and college major choice. This work is conducted within the context of an online learning environment of middle school mathematics used in classrooms, providing an opportunity to explore how data from such environments can be used to predict long-term educational outcomes – in the case of this dissertation research, intervention and support in keeping students on track towards the pathway to college.

Research Questions

This dissertation conducts three studies to answer the following research questions. Each study presents an analysis that uses different sets of data collected from overlapping sets of students, collected over the course of several years. The studies are as follows:

1. Are student behavior, academic emotions and knowledge during middle school computer-based math learning predictive of college enrollment and selectivity of the college attended? (Study 1)
2. Are student behavior, academic emotions and knowledge during middle school computer-based math learning predictive of the pursuit or choice of a STEM college major once students are in college? (Study 2)

3. How do high school course choices and interests in college majors and career during high school mediate between student behavior, academic emotions and knowledge in middle school computer-based math learning, and college attendance outcomes? (Study 3)

Study 1 assesses the behavior, emotions and knowledge of students who used a math educational software when they were in middle school, using their interaction data from the system. These middle school assessments were then integrated with more recent data on whether the students went to college, to create a structural equation model predicting long-term student outcomes (i.e. college enrollment, selectivity of college attended). Study 2 is similar to Study 1 but looks at how the middle school assessments can be predictive of a different college attendance outcome – what major students choose once they enroll in college. Study 3 first explores the relations between course choice and interest in college majors and careers during high school, the middle school assessments of student behavior, academic emotions and knowledge, as well as college outcomes. This analysis then leads to the development of an overall model that combines middle school assessments of student behavior, academic emotions and knowledge from computer-based math learning with their course choices, college and career interests when they were in high school, and with those students' college attendance outcomes. Study 3 examines how factors of choice and interest developed in high school can potentially mediate between the educational experiences students have during middle school (through assessments of student behavior, academic emotions and knowledge) and their eventual choices in going to college. Study 3 uses the same measures of student engagement and learning during middle school and data on college enrollment used in Study 1, coupled with survey data acquired when they were in high school about students' course choices, college and career interests. Study 1 and Study 2 use structural equation modeling and regression analyses to demonstrate how middle school factors are related to later college attendance outcomes. Study 3 uses regression analyses and mediational modeling to demonstrate how both middle school and high school factors are related to later college attendance outcomes. All of these models illustrate the cognitive and motivational mechanisms that students experience throughout varied phases in their years in school, and how they may be related to one another, providing implications for intervention designs (e.g. teacher reports) for educators.

This dissertation is divided into seven chapters. This first chapter has provided a background of the dissertation study, statement of the problem, study objectives, and research questions posed. Chapter Two presents a review of the literature related to the research questions and methods presented in this dissertation. Chapter Three describes the educational software, the ASSISTments system, used in this dissertation as a data source, and the related methodologies applied to extract information from the system relevant to this dissertation. Chapter Four examines the preliminary analyses conducted by the author as groundwork for the research questions in this dissertation. Chapter Five describes the data analyses and modeling conducted to address the research questions. Chapter Six presents the results and findings out of the analyses and modeling conducted. This dissertation concludes with a discussion of the results and their implications with regard to the research questions presented.

CHAPTER II.

REVIEW OF LITERATURE

Researchers in recent years have used educational technologies to explore constructs related to student learning, either in a laboratory or in actual classrooms. Computer-based learning environments provide a rich source of data that helps us understand students' learning processes. This data can help us model academic emotions and engagement. Academic emotions and engagement have been shown to influence cognition and deep learning, but have usually been investigated at a coarse-grained level (e.g. self-report measures, teacher ratings, interviews, observations), comparing them to performance in post-tests or end-of-year exams (Fredricks & McColskey, 2012; Wigfield, et al., 2008).

Within the context of online learning systems, recent studies have explored academic emotions, engagement, and learning in fine-grained detail, together with their associations with learning outcomes. Researchers have developed automated models that can infer students' academic emotions, engagement, and knowledge in real time, and have found evidence that the constructs these models infer are associated with differences in student outcomes. Specifically, these fine-grained assessments of cognitive and non-cognitive factors during middle school have been shown to predict learning gains (Baker, D'Mello, Rodrigo, & Graesser, 2010), performance on standardized exams (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013), and preparation for future learning (Baker, Gowda, Corbett, & Ocumpaugh, 2012; Hershkovitz, Baker, Gowda, & Corbett, 2013).

However, there has been limited research on whether these fine-grained measures can predict long-term student outcomes. This chapter discusses the possibility that these measures are related to and can be predictive of a long-term student outcome: the decision to pursue postsecondary education. The following sections establish the importance of the outcome of going to college and how student trajectories towards college attendance develop as early as middle school. Processes that influence decisions in pursuing college and careers are demonstrated through the Social Cognitive Career Theory (SCCT). This chapter then argues how fine-grained measures of educational constructs related to learning (i.e., academic emotions, engagement, knowledge) can be predictive of college attendance outcomes and how these measures can be factored into traditional models (i.e., SCCT). In addition to establishing and justifying these relations, this chapter demonstrates how measures of middle school constructs within the context of computer-based learning environments can be used for college and career counseling.

Importance of College and Postsecondary Education

Even with varying opinions about the role of postsecondary education in one's career trajectory, college enrollment and completion remain a necessary step towards career success (ACT, 2006; Carnevale & Desrochers, 2003). Today's modern economy is heavily dependent on a skilled labor force, and many jobs require qualifications that include a college degree. Employers typically require jobseekers to have a college degree and expect them to have training that equips them with the necessary skills for the job. This is especially important for industries based on science, technology, engineering, and mathematics (STEM; Smith, Morgan, & Schacht, 2003; Stine & Matthews, 2009).

It has been shown that getting a college degree is related to a higher chance of getting a job (Carnevale, Smith, & Strohl, 2010), higher levels of social and economic achievement (Haveman & Smeeding, 2006; Reardon, Baker, & Klasik, 2012), and better odds of improving one's economic status over time (U.S. Treasury Department, 2012). Individuals who have finished college earn substantially more than those who only finished high school (Gottschalk, 1997; Haveman & Smeeding, 2006; Hoxby, 2009; Lemieux, 2006). Many jobseekers without college degrees find themselves unable to secure stable employment, as job-generating industries demand qualified and trained workers. Hence, many argue that developing and improving the quality and size of the workforce starts with better educational preparation (Johnson, Nichols, Bubotz, & Riedesel, 2002; Rojewski, 2002; Sagen, Dallan, & Laverty, 2000). During this preparation, learners can be exposed to educational programs that may increase their interest and preparation for the careers they will choose. It is valuable for educators to evaluate the progress of students' trajectories toward successful entry into college. This includes assessing and developing the students' readiness and preparedness for college. Students transitioning to college are met with new academic and social environments and academic demands. Thus, educators must make students college-ready before they graduate high school by equipping them with the skills, interests, and information needed to succeed in college.

With the high need for workers with STEM training, it is important that educational programs and K-12 curricula do not just cater to already high-achieving students, but also to groups of students who may be interested in a particular career but lack the know-how to improve and develop their skills, to groups who are interested and engaged in domain-specific courses but are not properly guided on what potential career is suited for them, and to groups whose interest (and subsequent achievement) could be enhanced with appropriate scaffolding. These student groups are not always easy to identify when career guidance counselors assess students' vocational interest and career self-efficacy. The missed opportunity to support these groups of students can have a big impact once these students are finishing high school and considering their college options.

Gap between College-going Plans and Actual College Attendance

Even when students develop positive educational and career aspirations, there is still a disparity between their college-attendance plans and actual college attendance. Not all students get the opportunity to attend college once they finish high school. While students who developed their aptitude in middle or high school have a better chance of attending college (Christensen, Melder, & Weisbrod, 1975; Eccles, Vida, & Barber, 2004), very few high-achieving, low-income high school students apply to the best colleges in the United States (Hoxby & Avery, 2012). Many high-achieving but low-income students, despite their qualifications, end up in less selective or nonselective colleges (Bowen, Chingos, & McPherson, 2009) where they often do not graduate even though their high school records indicate that they are college-ready (Roderick, Nagaoka, & Coca, 2009). This phenomenon is particularly common among students whose parents do not have a college degree (Smith, Pender, & Howell, 2013). First-generation college students often do not see themselves as college-bound (Ishitani, 2006; Pascarella, Pierson, Wolniak, & Terenzini, 2004; Ramos-Sanchez & Nichols, 2007); more so, they usually have limited educational choices because of financial restrictions or obligations to their family (Inman & Mayes, 1999).

Students may find themselves in need of support to achieve their postsecondary plans. Students not continuing to college are diverse in aptitude and demographics, with access to college being skewed by students' race, ethnicity, and socioeconomic status. These students frequently receive less career guidance and counseling compared to those who plan to attend college (Herr & Niles, 1997). Minority students and students with lower socioeconomic status are reported to be least likely to seek support from academic or vocational counselors (Perrone, Sedlacek, & Alexander, 2001), and receive less concrete knowledge and support from their parents with regard to postsecondary planning (Valadez, 1998). These students often overestimate college costs, underestimate the availability of financial aid, and exhibit limited knowledge about academic prerequisites for college attendance (Avery & Kane, 2004). Specifically, White and Asian students are overrepresented and more likely to enroll in four-year colleges, whereas African American and Hispanic students are underrepresented (Carnevale & Rose, 2003; Reardon, Baker, & Klasik, 2012).

For these reasons, school counselors are encouraged to help students transition from secondary to postsecondary education (Gibbons, Borders, Wiles, Stephan, & Davis, 2006). They are encouraged to be aware of barriers that hinder students' school progress and create solutions to these issues (American School Counselor Association, 2005). It is important for school counselors to know about the specific needs of low-SES, minority, and first-generation students. This is vital in providing these students effective guidance and support in their college and career planning.

College Readiness and Current Assessments of College Success

There are many reasons for why some high school graduates are not college-ready. Conley (2010) identifies college readiness as the level of preparation a student needs to succeed without remediation at the postsecondary level. Both academic and nonacademic factors are relevant. While schools can influence guidance counselor practices or academic preparation, they cannot directly control factors stemming from family background, such as SES and parental educational background. Many studies over the past 10 years have documented the disconnect between what high school teachers teach and what postsecondary or college instructors expect, with regard to students' preparation for college. For instance, factors required for high school graduation, such as grade point average (GPA), class rank percentile, performance on standardized tests, and performance in college preparatory courses (e.g., advanced placement [AP] classes), have been found to be poor indicators of postsecondary outcomes, as high school graduates still find themselves in need of remedial courses upon entry to college (Conley, 2007, 2008, 2010). It is important to align expectations for high school graduation with college and career requirements, and important that students in develop core cognitive skills to be college-ready (Conley, 2008; Conley, Lombardi, Seburn, & McGaughy, 2009).

Equally important in preparing for college is the development of non-cognitive skills. These non-cognitive skills include behavioral, emotional, and attitudinal factors that allow students to successfully manage new learning content, environments, and demands (Dowson & McInerney, 2003; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Empirical studies have examined cognitive and non-cognitive factors that influence a student's probability of college-attendance, using students who are already in college. However, using study participants at this stage can overlook what led to students not pursuing college. Hence, it can be argued that research on why students pursue postsecondary education should examine the period before entry to college, and evaluate factors in the students' decision-making process on whether to enroll in college. While demographic information and academic standing are important for college and career counseling, they do not illuminate all the possible reasons why students fail to attend college. Gibbons et al. (2006) identified that beyond accurate knowledge about college costs, assessment of students' academic and career self-efficacy may be valuable information in counseling efforts. Farrington and colleagues (2012) identified non-cognitive factors such as academic behaviors, academic perseverance, academic mindsets, learning strategies, and social skills as influential to students' long-term success. They argued that such factors are receptive to contextual influences. Hence, understanding students' long-term outcomes such as college readiness, attendance, and persistence necessitates looking beyond their academic performance and individual abilities. Counseling efforts should also consider the students' experiences within their educational environments, assessing their behaviors, attitudes, and motivation during learning.

The Social Cognitive Career Theory

The factors that come into play between the students' environment and their learning experiences can be seen within the Social Cognitive Career Theory (SCCT; Lent, Brown, & Hackett, 1994, 2000). SCCT emphasizes the interplay between environmental and individual factors that contribute to academic and career choices students make (Lent & Brown, 2006). Based on Bandura's (1986) general social cognitive theory, SCCT asserts that academic and career choices are shaped throughout middle school and high school by constructs such as environmental supports and barriers, as well as the students' self-efficacy, outcome expectations, goals, and interests (see Figure 1). This means that activities that contribute to positive experiences and higher self-efficacy in students help form their interests and engagement in those activities. Conversely, students avoid and become less interested in activities that lead to negative outcomes and a decrease of self-efficacy.

According to SCCT, environmental supports and barriers play a significant role in influencing choices. Factors related to ethnicity, race, socioeconomic status, family background, and gender, may create negative outcome expectations. The effects of these environmental factors are also evident when students transition to college and are faced with a new series of demands (e.g., financial resources, academic integration to college) (Wang, 2013). Hence, academic and career counselors must help students think about these factors and advise them on ways to cope during this transition. Turner and Lapan (2002) examined the contributions of students' career self-efficacy, planning, and exploration to the formation of their career interests. They showed that perceived parental support influences middle school students' self-efficacy, which then influences their career interests (Turner & Lapan, 2002).

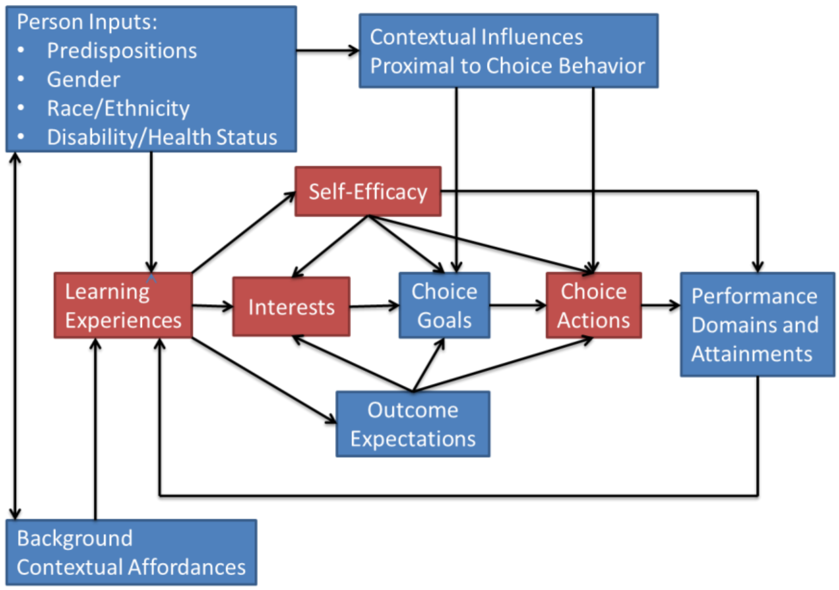


Figure 1. Social Cognitive Career Theory (SCCT).

As mentioned, SCCT posits that higher levels of interest emerge in contexts where the individual has higher expectations of self-efficacy and outcome expectations, with these interests leading to the development of intentions or goals for further exposure and engagement with the activity (Lent, Brown, & Hackett, 1994). Self-efficacy and outcome expectations can then serve either as supports or barriers to students' long-term success in college and their careers. In general, self-efficacy is related to four factors (Bandura, 1977): experiences of achievement; vicarious learning; persuasion through encouragement or discouragement; and emotional, behavioral, or physiological states (e.g., anxiety, self-esteem, etc.). In relation to careers, self-efficacy refers to an individual's perception and belief about career-related behaviors that influence educational or occupational choices and participation in those choices (Betz & Hackett, 1997). This may include engagement in technical courses related to a specific career or after-school programs that teach vocational skills. Hence, it can be posited that interest mediates between self-efficacy and student choices, with self-efficacy mediating between student performance and the formation of interests. These associations suggest that meaningful and effective learning can increase self-efficacy and in turn influence interest formation. It is thus important to identify factors that govern students' learning experiences prior to making choices related to college outcomes, and evaluate how these experiences contribute to their self-efficacy and interest formation.

SCCT Factors during Middle School

While research in SCCT and college and career readiness has usually focused on high school or college students (Byars-Winston & Fouad, 2008; Gore, 2006; Wang, 2013), this dissertation explores these phenomena during middle school as well. Relatively few studies (Fouad & Smith, 1996; Gibbons & Borders, 2010; Turner & Lapan, 2002) have analyzed hypotheses stemming from SCCT during middle school. This is a surprising exclusion since middle school has been found to be a key phase in students' development of abilities and interests that impact their pursuit of postsecondary education and careers (Cabrera, La Nasa, & Burkum, 2001; Camblin, 2003).

In particular, college planning occurs during middle school and high school, with the U.S. Department of Education recommending college planning as early as sixth grade (US Department of Education, 1999). Research suggests that students' vocational interests can be fairly stable as early as middle school, though students continue to explore college and career options during high school (cf. Blustein, 1992; Lent, Hackett, & Brown, 1999; Tracey, Robbins, & Hofsess, 2005). During middle school, students begin to develop academic abilities, interests, and choices that will have a strong influence on later academic outcomes (Cupani & Pautassi, 2013). Middle school students become engaged or disengaged from school and learning, driven in part by changes in self-perception such as whether they see themselves as smart and capable of going through high school. Students who start thinking about college as early as middle school tend to become interested in achieving a good academic record. They may plan to take appropriate courses once they are in high school or choose to be involved in extracurricular activities that will contribute to their college applications (Roderick, Coca, & Nagaoka, 2011; Roderick, Nagaoka, Coca, & Moeller, 2008).

Conversely, changes in terms of greater academic failure or decreasing motivation also begin to manifest in middle school (Anderman & Maehr, 1994; Neild, 2009; National Middle School Association, 2002). Many students drop out of the pipeline to academic success well before they start thinking about college (Balfanz, 2009; Balfanz, Herzog, & Mac Iver, 2007; Bowers, 2010; Bowers & Sprott, 2012). Many of them exhibit problem behaviors and extreme forms of disengaged behavior, such as low attendance and misconduct, manifested in disciplinary referrals (Tobin & Sugai, 1999; Tobin, Sugai, & Colvin, 1996). If these changes could be spotted early, better interventions could be developed to support these students (Bowers, 2010). In addition, by the time these fairly strong indicators of disengagement are known, it may be quite late to intervene. If it were possible to identify useful, alterable, or actionable antecedents to these changes, it might be possible to intervene more effectively.

Christenson and Thurlow (2004) have suggested that interventions should emphasize school completion rather than dropout. School completion requires a focus on student behaviors and attitudes compatible with the school's practices and expectations. Success can be supported by educators focusing on developing students' competencies rather than attempting to address their deficits, and by tailoring interventions to fit individual students (Christenson & Thurlow, 2004). An intervention focus on student engagement involves formulating ways to increase students' engagement, motivation, and interest to learn. A student who may be performing well in a specific domain or subject (i.e., math, science, art) but who is not currently interested in pursuing that domain in college or as a career (and vice versa) may be a target of interventions and programs geared toward interest development for a certain domain. Thus, measurable and actionable factors that support a student's plan to attend college must be identified as early as middle school. Both cognitive and non-cognitive aspects of students' educational experiences, such as student learning, academic emotions, and behaviors during middle school, can be malleable and actionable. These factors influence vocational interest and self-efficacy for a particular career. For example, differences in learning influence a student's self-efficacy for a particular domain, with poor learning reducing self-efficacy whereas successful learning increases self-efficacy (cf. Bandura, 1997). Linnenbrink and Pintrich (2003) frames student engagement, learning and achievement as reciprocally related to self-efficacy, where self-efficacy leads to more engagement and subsequently more learning and better achievement, and in turn increases self-efficacy. Student engagement in this context is broken down into different aspects such as behavioral engagement that includes the observable behavior of students with respect to their effort, persistence and help-seeking; cognitive engagement that is related to the students' active learning and the learning strategies that they employ such as self-regulation and metacognitive strategies; and motivational engagement that pertains to the students' displays of interest and value in their learning activities that contribute to their affective state or academic emotions during those learning tasks (Linnenbrink & Pintrich, 2003). Thus, it can be argued that factors such as knowing a skill, academic emotions and student engagement during middle school may play an influential role in the processes involved in SCCT, and therefore may contribute to the eventual decision to attend college. In SCCT, students' initial vocational interests are modified by their self-efficacy, attitudes, and goals for career development (i.e., college enrollment, career interest, STEM interest), which can be seen as themselves influenced by students' engagement when they encounter increasingly sophisticated domain content (see examples in Baker, 2007; Baker et al., 2008; McQuiggan, Mott, & Lester, 2008).

Fine-Grained Assessments with Educational Technology

Academic emotions, engagement, and learning during middle school have usually been investigated at a coarse-grained level and their association with long-term student outcomes has only infrequently been studied. However, educational technologies have been used by researchers in recent years to explore the relationship between these constructs and eventual student outcomes. These systems offer large-scale data sets with high-quality, fine-grained interaction data. For example, the ASSISTment system (Razzaq et al., 2005) was used by over 50,000 students in the Northeastern United States in 2012–2013 as part of their regular middle school mathematics classes. Other systems such as ALEKS (Canfield, 2001) and the Cognitive Tutor (Koedinger & Corbett, 2006) are used by hundreds of thousands of students each year. Through these systems, students produce a series of actions as they complete learning activities, yielding a rich source of data that can support researchers in investigating whether students' strategic choices and behaviors translate into learning, providing the potential for rich, multi-faceted, and fine-grained assessments of these constructs (Clarke-Midura & Dede, 2010).

Educational Data Mining for Fine-Grained Assessments

The growth of new technologies used in the educational context (e.g. online learning, educational games, learning management systems) has led to the increasing quantity of data captured that can be used for research in exploring patterns of different educational phenomena.

Processes in extracting useful and actionable information from such large databases or datasets used in analyses have adopted methodologies from data mining, machine learning, data visualization, psychometrics and other areas of statistics (Baker & Yacef, 2009). The application of these varied analytics methods to education has formed the discipline of educational data mining (EDM) that takes an *a posteriori* approach to data, where data (usually from large data sets) is analyzed retrospectively to discover patterns that may be overlooked when testing pre-determined hypotheses (common in most traditional statistical approaches). Unlike most of the data mining methods used in other domains, EDM exploits the meaningful hierarchies that can be inherent in educational data (e.g. district level, school level, student level, etc.). Like traditional statistics, EDM is concerned with discovering structure in data by creating models or exploring relationships between variables. But unlike traditional statistics, EDM also uses the models created to discover meaningful patterns and then use them for prediction on new datasets. In the educational setting, predictive modeling is often used to make operational decisions to improve educational outcomes (e.g. academic performance, attendance, graduation rates, etc.). For example, predictive modeling in education can be used to identify students who are at risk, predict student performance, predict on-time graduation, examine indicators of readiness for college and career, or personalize instruction in classrooms. With its roots in analyzing student-computer interaction (i.e. educational software), one main characteristic of EDM methods involves the automated discovery of constructs or patterns within educational data that can be used for adaptation and personalization within systems (Baker & Siemens, 2014). EDM methods have been used in modeling student individual differences in areas such as student knowledge, motivation, and meta-cognition to enable systems to respond to these differences and improve student learning (Baker & Yacef, 2009).

In recent years, EDM researchers have modeled a range of student attributes and examined the relationships between them, both within educational software (e.g. student academic emotions and behavior) and beyond the context of the educational software (e.g. performance in state exams). Using logs of student interaction with these systems, researchers in the student modeling and educational data mining communities have developed automated models that can infer students' academic emotions, engagement, and knowledge in real time (Baker et al., 2008; Corbett & Anderson, 1995; Razzaq et al., 2005). Recent advances in student modeling (Desmarais & Baker, 2012) and educational data mining (Baker & Yacef, 2009) have resulted in fine-grained measures of these constructs and evidence on how they relate to student outcomes. These models, often developed from a combination of expert field observation (e.g., Ocumpaugh, Baker, & Rodrigo, 2012) and data mining on interaction logs, can accurately predict expert labels of academic emotions and engagement on entirely new students (cf. Baker, 2007; Baker, Corbett, & Aleven, 2008; Baker, Corbett, Koedinger, & Wagner, 2004; Baker et al., 2012; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Walonoski & Heffernan, 2006). EDM researchers typically develop a model of a construct within an online learning environment by using a multi-step process that leverages ground truth labels of this construct usually obtained from human judgments. These labels are used to train a model of the construct so that it can be used to detect or infer the occurrence of this construct when human judgments are not feasible.

As will be shown in this dissertation study, models inferring learning (i.e., knowledge), academic emotions and behavior, to be discussed in a later section, were applied to the interaction data of this study's sample population, creating features or attributes that will be used for the outcome models predicting college attendance and other variables. This EDM approach is sometimes referred to as "discovery with models" (e.g. Hershkovitz, Baker, Gobert, Wixon, &

Sao Pedro, 2013), where existing models, derived from student modeling and machine learning methods, are used as a component in a new and different analysis or model. Assessments or measures from these models are different from the questionnaire responses and coarse-grained measures typically used in educational research. Assessments developed using student modeling/machine learning have been shown to predict educational outcomes such as learning gains (Baker, Corbett, Koedinger, & Wagner, 2004; Cocea, Hershkovitz, & Baker, 2009; Sabourin, Mott, & Lester, 2011) and standardized exams (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013), and have been widely used in recent years in studying educational phenomena within the context of online learning environments that produce rich student interaction data such as intelligent tutoring systems (Baker, D'Mello, Rodrigo, & Graesser, 2010; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Walonoski & Heffernan, 2006) and educational games (Shute, Moore, & Wang, 2015; Bosch et al., 2015).

With the potential for evaluating student outcomes, fine-grained measures of constructs in the students' learning experiences may be able to predict eventual long-term outcomes such as college enrollment, selectivity of college attended, or choosing a particular college major, while also providing the potential for immediate action.

Cognitive and Non-Cognitive Factors in Online Learning Environments

Recent studies in online learning systems such as tutoring systems and educational games have explored fine-grained measurements of cognitive and non-cognitive factors during a student's interaction and learning experience with those systems. As previously mentioned, models that can infer students' knowledge of a certain skill, academic emotions, engaged and disengaged behaviors in real time have been developed within these environments to obtain these fine-grained measurements (Baker et al., 2008; Corbett & Anderson, 1995; Razzaq et al., 2005).

Details of how these constructs can be modeled will be discussed in the succeeding chapter for a particular instance of a learning system.

Student knowledge is estimated during a student's interaction with learning systems by modeling how much a student knows a required skill whenever the student goes through a learning task within a system (i.e. giving an answer, using a hint) (Corbett & Anderson, 1995). Academic emotions and behaviors of engagement and disengagement that have been studied within learning systems include those which are prominent in traditional classroom settings and widely known to influence on cognition and learning outcomes (Baker, D'Mello, Rodrigo, & Grasser, 2010; Baker et al., 2012; D'Mello, Taylor & Grasser, 2007; Dragon et al., 2008; Lee, Rodrigo, Baker, Sugay, & Coronel, 2011; Sabourin, Rowe, Mott, & Lester, 2011).

Academic emotions are common and can play an important role in learning outcomes, and have also been shown to be measurable and actionable antecedents of engaged and disengaged behaviors during learning (Baker, D'Mello, Rodrigo, & Graesser, 2010; D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; D'Mello & Graesser, 2012; Pekrun, Goetz, Titz, & Perry, 2002; Pekrun & Linnenbrink-Garcia, 2012; Rowe, McQuiggan, Robison, & Lester, 2009). Academic emotions are different from Ekman and Friesen's (1978) basic emotions in everyday experience: fear, anger, happiness, sadness, disgust, and surprise. Academic emotions, also referred to by some as affective states, are more specifically the emotions that are relevant in educational settings, influencing cognition and deep learning (Kort, Reilly, & Picard, 2001; Craig, Graesser, Sullins, & Gholson, 2004).

An example is *boredom*, which is prominent in many middle school classrooms (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010; Rowe, McQuiggan, Robison, & Lester, 2009). A classroom session may be received by a student with a level of interest and enthusiasm at the beginning. However, as the session progresses, boredom may set in when novelty of the content and learning environment fades, or when students have difficulty comprehending the lesson.

A second affective state, *engaged concentration*, is related to Csikszentmihalyi's flow state (1990); it describes the state when a student has intense concentration, focused attention, and complete involvement in the task at hand (Baker, D'Mello, Rodrigo, & Graesser, 2010). This affective state is differentiated from Csikszentmihalyi's construct of flow by lacking its task-related aspects such as clear goals, immediate feedback, and balance between challenge and skill (e.g., a student can experience engaged concentration even if the challenge is significantly higher than their skill).

Another academic emotion is *confusion*, where a student encounters a mismatch in their understanding that is not immediately resolved between their prior knowledge and incoming information, creating a cognitive disequilibrium in students (D'Mello, Lehman, Pekrun, & Graesser, 2014; Rozin & Cohen, 2003). Researchers have investigated how common this affective state is in complex learning and how it relates to learning outcomes. Craig and colleagues (2004) found that confusion was positively associated with learning gains and engaged concentration, while Rodrigo and colleagues (2009) found confusion to be negatively associated with achievement. However, prolonged, unresolved confusion is associated with poorer student performance (Lee, Rodrigo, Baker, Sugay, & Coronel, 2011; Liu, Pataranutaporn, Ocumpaugh, & Baker, 2013).

Students can also experience *frustration* (Kort, Reilly, & Picard, 2001) which, like confusion, promotes cognitive disequilibrium in students. With frustration, students have feelings of distress when they encounter tasks that may be too difficult for their skills

(Csikszentmihalyi, 1990). Frustration can also be a natural part of a student's cognitive processing; in many cases, frustration may not need external intervention and can eventually be resolved (Mentis, 2007). Like confusion, frustration is associated with poor learning when it is prolonged and unresolved, but it can also be associated with learning gains when it occurs only briefly (Liu, Pataranutaporn, Ocumpaugh, & Baker, 2013).

Negative academic emotions can lead students to zone out (Drummond & Litman, 2010; Feng, D'Mello, & Graesser, 2013) or exhibit disengagement in classrooms. Examples of disengaged behaviors during learning can include gaming the system, off-task behavior, and carelessness. *Gaming the system* is a behavior when a student exploits the properties of a learning activity (i.e., within an educational software) to obtain the solution instead of through meaningful learning (Baker, Corbett, Koedinger, & Wagner, 2004). It includes systematic guessing and using hints or help features to get the answers. A second relevant disengaged behavior is off-task behavior. When students exhibit off-task behavior, they engage in extraneous activities and completely disengage from their learning tasks. *Off-task behavior* has been documented in both computer-supported and traditional learning activities (Karweit & Slavin, 1982; Baker, Corbett, Koedinger, & Wagner, 2004). There are several manifestations of off-task behavior, including talking to a classmate, passing notes, or surfing the Web. Within the context of educational software, off-task behavior has been associated with poorer learning (Baker, 2007; Cocea, Hershkovitz, & Baker, 2009; Rowe, McQuiggan, Robison, & Lester, 2009). In learning activities, students have also been found to exhibit *careless behavior* when they make errors on questions despite knowing how to successfully answer them (Clements, 1982). This appears to be a common occurrence when students use educational software, such as the Cognitive Tutor (San Pedro, Baker, & Rodrigo, 2011).

These disengaged behaviors, together with the affective state of boredom, have been found to lead to poorer learning, lower self-efficacy, diminished interest in educational activities, and, most importantly, increased attrition and dropout rates (Baker, D'Mello, Rodrigo, & Graesser, 2010; Craig, Graesser, Sullins, & Gholson, 2004; Daniels et al., 2009; Goodman, 1990; Mann & Robinson, 2009; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010; Wasson, 1981). Gaming the system has been associated with negative attitudes toward math content (Baker et al., 2008), poorer performance on end-of-year exams (Baker, Corbett, Koedinger, & Wagner, 2004), and poorer learning compared to students who do not game the system (Cocea, Hershkovitz, & Baker, 2009). Like gaming the system, off-task behavior is related to students' negative attitudes toward math content (Baker, 2007) and lower self-efficacy (Narciss, 2004; Schunk, 1989).

At the other end of the spectrum are students who are more engaged in school and tend to have higher academic motivation and achievement (Fredericks, Blumenfeld, & Paris, 2004; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013). Research studies on the relationships between academic emotions and learning have found that engaged concentration is positively associated with learning outcomes (Craig, Graesser, Sullins, & Gholson, 2004; Csikszentmihalyi, 1990; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Rodrigo et al., 2009).

As established in this chapter, academic emotions and student behavior are likely to play an important role in the development of academic and career self-efficacy and interests, and they can be indicators of being prepared for college. These factors can thus serve as additional information and predictors in current models for college and career pathways. As in SCCT, student knowledge, academic emotions and behavior that contribute to a student's learning experiences that can be emphasized as crucial factors for support and guidance (Figure 2).

Beyond this, they can be studied in terms of how they drive and interact with other instructional and motivational processes that lead to students' college and career choices. This richer information can also be included in reports (e.g. in software dashboards) that may assist educators in identifying at-risk students and encourage those students to participate in educational activities and programs tailored to their specific learning needs, so as to keep them in the academic pipeline.

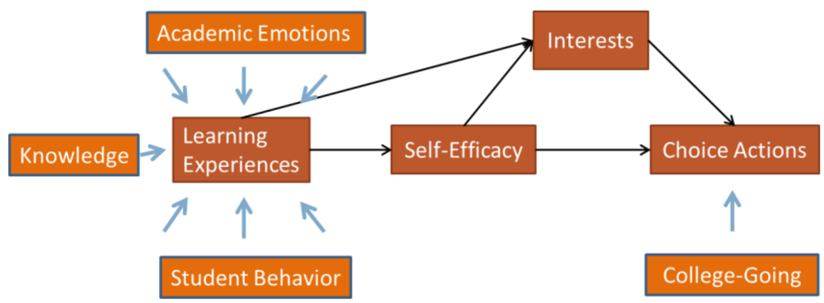


Figure 2. SCCT Model with Constructs of Knowledge, Academic Emotions and Student Behavior.

CHAPTER III.

FINE-GRAINED MODELING IN THE ASSISTMENTS SYSTEM

To address the research questions for this dissertation, the author leveraged interaction data from the ASSISTments system and its fine-grained measures of cognitive and non-cognitive factors during middle school math learning. A range of constructs were assessed from this interaction data including student knowledge estimates, student academic emotions (boredom, engaged concentration, confusion, frustration), student disengaged behaviors (off-task, gaming the system, carelessness), and other information on student usage (the proportion of correct actions and the number of actions – a proxy for overall usage), to form the variables used in creating the outcome models. These variables were either directly obtained from the interaction data or from classifications or assessments from models applied to the interaction data.

This chapter discusses the online learning environment that is the primary source of middle school data – the ASSISTments system – and how models of student knowledge, academic emotions and student behavior were developed and applied in the ASSISTments system. In particular, this chapter details how models/detectors of boredom, engaged concentration, confusion, frustration, off-task behavior and gaming the system for ASSISTments in (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013) and in (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014) were created from representative students, validated, and applied to the interaction data of a different student sample used in this dissertation. In addition, this chapter also discusses the models of student knowledge and carelessness used for the dissertation's student sample.

The ASSISTments System

The ASSISTments system (Figure 3) is a tutoring system for middle school mathematics provided by Worcester Polytechnic Institute (WPI) (Razzaq et al., 2005). This free web-based educational system aims to *assess* knowledge and proficiency of its student users while *assisting* them in their problem solving and learning. ASSIStments provides teachers with detailed reports and summaries on the mathematical skills each student learns. The system delivers mathematics problems and questions, assesses student performance, provides hints and suggestions, provides targeted feedback on common errors, and scaffolds the development of improved answers by breaking complex problems into simpler steps. Within the system, each mathematics problem maps to one or more knowledge components or mathematical skills. These knowledge components or skills cover a range of areas in mathematics, including algebra, probability, number sense with fractions and decimals, geometry, and graph interpretation. When students working on an ASSISTments problem answer correctly, they proceed to the next problem. If they answer incorrectly, they are provided with scaffolding questions where the problem is broken down into its component steps in order to concretize the systematic thinking needed to solve the problem. The intention for this is to identify which part of the student's thinking is incorrect. Each step of the scaffolding, which involves either the same or different math skill as the original problem, is also a problem requiring a new answer, with its own set of hints. The last step of scaffolding returns the student to the original question (as in Figure 4). Once the correct answer to the original question is provided, the student is prompted to go to the next question. In this way, the students learn mathematics while the system learns which steps the students could not do without assistance. This information about the student's problem solving is then provided to teachers for assessment and diagnostic purposes.

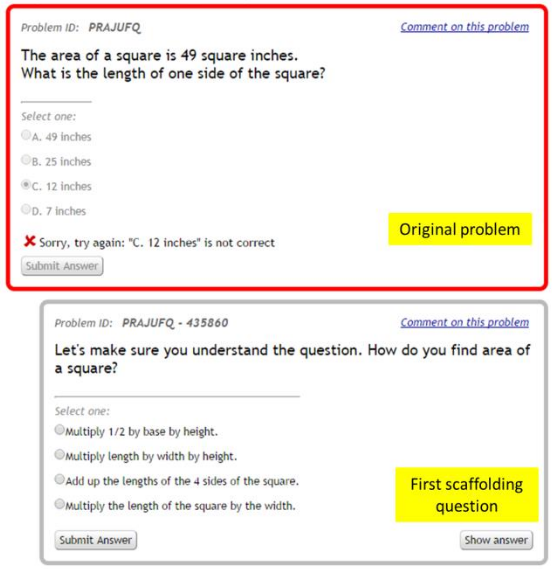


Figure 3. Example of an ASSISTments problem.

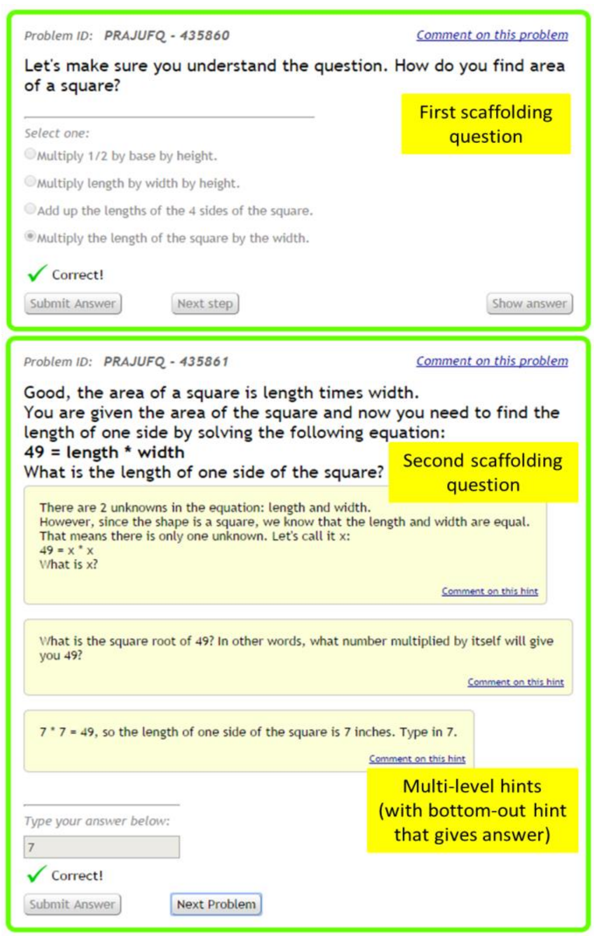


Figure 4. Example of scaffolding and hints in an ASSISTments problem.

Modeling Academic Emotions, Behavior, Student Knowledge for ASSISTments

Models inferring learning (i.e., knowledge), academic emotions and behavior were applied to the interaction data of the student sample in this dissertation, creating features, attributes or variables that will be used for the outcome models predicting college attendance. The models of boredom, engaged concentration, confusion, frustration, off-task behavior and gaming the system – to be discussed later – were created by first labeling student academic emotions and engagement (through classroom field observations) from a small but reasonably representative sample of students who used the educational software, and synchronizing these labels with the interaction data generated by the software during their usage to create the training data to generate the models of academic emotions and engagement (Figure 5). These models were then applied to interaction data at scale – data from a different larger sample of students who used the software (in this dissertation study, the student sample), to then produce their measures of academic emotions and behavior.

These models for the ASSISTments system were developed and first used in (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013) to assess the relationship of academic emotions and behavior and math state test scores, and validated more thoroughly in (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014), which assessed the models' validity across multiple populations. This dissertation applies these models of boredom, engaged concentration, confusion, frustration, off-task behavior, and gaming the system to the interaction data for this dissertation's student sample.

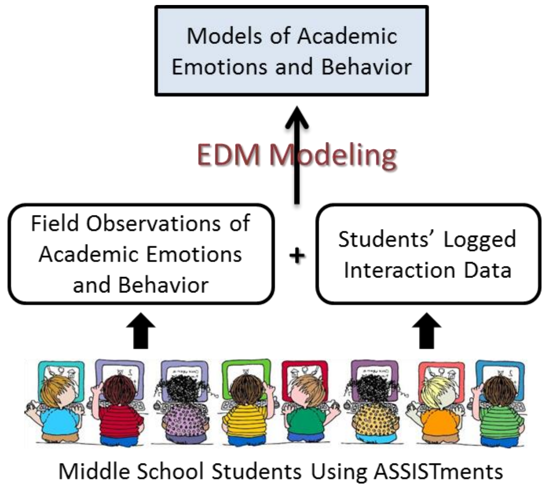


Figure 5. Modeling academic emotions and behavior in ASSISTments.

Student knowledge was assessed on the interaction data of the student sample for this dissertation using a model that generates Bayesian inference from the student's correct and incorrect responses to a problem step that is associated with a skill (Bayesian Knowledge Tracing, Corbett & Anderson, 1995). Carelessness, while a form of student behavior, was similarly assessed with a model created from Bayesian inferences (to be discussed below).

Academic Emotions and Disengaged Behavior in ASSISTments

For student academic emotions (or affect/affective states) and behavior features, assessments of these constructs were obtained by utilizing existing models of academic emotions and behaviors previously developed for the ASSISTments system (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014), to help us understand student academic emotions and behavior across contexts. The academic emotions modeled within ASSISTments consist of boredom, engaged concentration, confusion, and frustration. Student disengaged behaviors modeled consist of gaming the system and off-task behavior. The resulting interaction data include a sequence of predictions of students' academic emotions and behavior across the history of each student's use of ASSISTments.

For the ASSISTment system, three separate models were developed for each academic emotion – one for students in urban schools, one for students in suburban schools, and one for students in rural schools. This is based on evidence that urban, suburban, and rural students manifest their emotions differently in online learning (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014). For off-task behaviors off-and gaming the system, only one set of models (urban) was developed – the models were trained just on urban students; these models were reused for suburban students, as they were found to be valid for this population as well (performing equally effectively when applied to new students from the different population, as when applied to new students from the original training population) (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014).

For the student sample in this dissertation, the urban set of models (Pardos, et al., 2013; Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014) were used to measure academic emotions within the interaction data of students who attended urban schools, while the suburban set of models (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014) were used to measure academic emotions within the interaction data of students who attended suburban schools. For gaming the system and for off-task behavior, interaction data from students who attended either urban or suburban schools were assessed with the (original) urban set of models, based on evidence of validity for both data sets.

The process for developing sensor-free models of academic emotions and student behavior for ASSISTments in both urban and suburban sets replicated a process which was previously successful for developing models of academic emotions or affect detectors for the Cognitive Tutor Algebra (Baker et al., 2012), and subsequently for other educational systems as well, such as Reasoning Mind (Miller, Baker, Labrum, Petsche, & Wagner, 2014) and Physics

Playground (Kai et al., 2015). These models were developed using a three-stage process: first, field observers coded student engagement and academic emotions using the BROMP protocol for quantitative field observation of emotion and engagement (Ocumpaugh, Baker, & Rodrigo, 2012) and the HART field observation app for Android (Baker et al., 2012) while students used ASSISTments; second, those field notes were synchronized with the interaction data from ASSISTments at a precision of around a 1-2 second error, using an internet time server; and third, data mining techniques were used to create models that could predict the field observations (i.e. student academic emotions and behavior) from the interaction or log data.

For the urban set of models, field observations of academic emotions and behavior were conducted in an urban middle school in New England, sampled from a diverse population of 229 students. These observations served as ground truth labels for boredom, engaged concentration, confusion, frustration, off-task behavior, and gaming the system. Within this school, the population included comparable proportions of Hispanic, African-American and Caucasian students, with per capita income significantly lower than the state average. For the suburban set of models, field observations for academic emotions and behavior were conducted in three suburban schools in New England, sampled from a total of 243 students predominantly comprised of White and East-Asian students of mid-to-high socioeconomic status, with less than 20% of students receiving a free or reduced-price lunch.

With the BROMP method, academic emotions and behavior were coded by a pair of expert field observers as students used ASSISTments. Each observation lasted up to twenty seconds, with elapsed observation time so far displayed by the hand-held observation software. If it was possible to label the academic emotion or behavior before twenty seconds elapsed, the coder moved to the next observation. Each observation was conducted using side glances, to reduce observer effects. The observers based their judgment of a student's academic emotion on the student's work context, actions, utterances, facial expressions, body language, and interactions with teachers or fellow students. These are, broadly, the same types of information used in previous methods for coding academic emotions (e.g., Bartel & Saavedra, 2000), and in line with Planalp et al.'s (1996) descriptive research on how humans generally identify affect using multiple cues in concert for maximum accuracy rather than attempting to select individual cues. At the beginning of data collection, an inter-rater reliability session was conducted, where the two coders coded the same student at the same time, 51 times. The resulting inter-reliability from this session was acceptably high, with Cohen's Kappa of 0.72 for categories of academic emotions (agreement 72% better than chance), and a Cohen's Kappa of 0.86 for categories of student behavior (agreement 86% better than chance).

Both the handhelds and the educational software logging server were synchronized to the same internet time server during observations, allowing logged student actions to be precisely correlated to the observations. The original log files consisted of data on every student attempt to respond (and whether it was correct), and requests for hint and scaffolding, as well as the context and time taken for each of these actions. Interactions with the software during the twenty seconds prior to data entry by the observer were aggregated into a clip, and data features were distilled.

The models were constructed using only log data from student actions within the software occurring at the same time as or before the observations, making the models usable for real-time automated interventions, as well as the discovery with models analyses here. Each of the models of academic emotions and behaviors used combinations of features engineered from raw information about a student's interaction (e.g. action is a hint, first attempt at a problem is a help request, etc.) to make predictions of that emotion or behavior, discussed below. Common classification algorithms in educational data mining were used in modeling each construct for this research, using the model with the best performance. These algorithms included J48 decision trees, logistic regression, JRip, Naïve Bayes, REP-Trees, and K-Star (Witten & Frank, 2005). The J48 classifier builds a C4.5 decision tree (Quinlan, 1992) from a set of labeled training data. The algorithm splits the data samples into smaller sample subsets based on an attribute that is most useful in discriminating between the classes to be learned (i.e. information gain), repeating this process on those smaller subsets until a decision node can be created that chooses a class. Logistic regression predicts the probability of a binomial outcome based on the use of one or several predictors. Logistic regression is similar to a linear regression, but rather than the probability, the curve is built using the natural logarithm of the “odds” of the outcome variable resulting to predicted values between 0 and 1. The JRip algorithm learns if-then rules that are easy to interpret. It generates the default rule first and then the exceptions for the default rule with the least (weighted) error rate. Naive Bayes classifier applies a simplified version of Bayes rule in order to compute the posterior probability of a category given the input attribute values of an instance, whose prior probabilities are estimated from frequency counts computed from the training data. The REP (Reduces Error Pruning) tree classifier applies regression tree logic and generates multiple trees in altered iterations using variance and information gain. K-Star is an instance-based classifier that predicts the class of a test instance based upon the class of those training instances similar to it, as determined by a similarity function.

Each of these models were cross-validated by repeatedly building them on in-sample data (also called training data) composed of a subset of the available data (4/5 of the 229 urban students; 4/5 of 243 suburban students), and testing them on out-of-sample data (also called test data) – the other 1/5 of the students – using the goodness metric A' to select the best model for each construct (shown in Table 1). The A' metric assesses each model’s confidence in classifying an emotion or behavior. This metric indicates the probability or percentage of time that given a single positive example and a single negative example, the model will accurately identify which is which. For example, the gaming model had an A' of 0.802, so the gaming model could distinguish a gaming student from a non-gaming student 80.2% of the time. An A' value of 0.5 indicates chance-level performance, and 1.0 indicates the model performs perfectly.

The A' metric closely approximates the area under the ROC (Receiver-Operating Characteristic) curve, called AUC (Hanley & MacNeil, 1982). The ROC curve describes the relationship between the true positive ratio and the false positive ratio predicted by a model, while the AUC represents the probability of a model being able to identify a randomly selected positive sample from a randomly selected negative sample across all probability thresholds for distinguishing a positive sample from a negative sample (Hanley and McNeil, 1982). A’ is often calculated as AUC, but a lot of existing statistical packages have inflated AUC computations, inaccurately measuring it in special cases where the data is skewed. A more accurate A’ implementation uses a comparison function that assigns a score (i.e. 0, 0.5, or 1) when comparing the model predictions for an observed positive and an observed negative sample, for every pair of positive and negative samples (Fogarty, Baker, & Hudson, 2005).

With AUC being analyzed in terms of A’, Hanley and MacNeil (1982) also shows that A’ is mathematically equivalent to the Wilcoxon statistic. This makes the A’ metric useful in conducting statistical tests on model comparisons (i.e. whether A’ values are significantly different between two or more models, or different datasets – a model with A’ of 0.83 is always better than a model with A’ of 0.80), or whether a model is significantly better than chance. Compared to other metrics of model goodness, such as accuracy, Cohen’s kappa or the F-1 score, A’ or area under the ROC curve is more robust in situations of imbalanced class distributions or skewness (Jeni, Cohn, & De La Torre, 2013).

Table 1

Model Performances (A’) of Urban and Suburban Models of Academic Emotions and Behaviors

	Boredom	*Engaged Concentration*	*Confusion*	*Frustration*	*Off-Task*	*Gaming*
Urban Model A’	0.632	0.678	0.736	0.743	0.819	0.802
Suburban Model A’	0.666	0.631	0.744	0.589	N/A	N/A

The best boredom model for students from urban schools was found using the JRip algorithm achieving an A' of 0.632, while the best boredom model for students from suburban schools used the REP-Tree algorithm with an A' of 0.666. The best model of engaged concentration for students from urban schools involved the K-Star algorithm, with an A' of 0.678, while the model for students from suburban schools used the J48 algorithm with an A' of 0.631. The best confusion model for students from urban schools used the J48 algorithm with an A’ of 0.736, and the best confusion model for students from suburban schools used the REP-tree algorithm achieving an A' of 0.744. The best frustration model for students from urban schools achieved an A' of 0.743 using the REP-Tree algorithm, and the best frustration model for students from suburban schools also used the REP-Tree algorithm with an A' of 0.589. The best model of off-task behavior used for students from both urban and suburban schools was found using the REP-Tree algorithm, with an A’ value of 0.819. Lastly, the best gaming model used for students from both urban and suburban schools involved the K-Star algorithm, having an A’ value of 0.802. This entire process resulted in automated models of academic emotions and engagement that can be applied to interaction data at scale, specifically log data of different students from the same learning environment, such as the data set used in this dissertation.

The patterns identified by each of these models are complex (see Appendix A for full detailed models, except for K-Star models that do not have an output model from the data mining package used). However, some of the core behaviors identified by each model are provided below. The boredom model trained for students from urban middle schools deems students to be bored based largely on lengthy pauses while using the tutor, and working on the same problem for some time but still not getting it correct (a serious and actively working student will generally obtain some correct answers in ASSISTments, as increasingly easy scaffolding is given when students make errors). For suburban students, boredom detection largely identifies students as bored based on tutor usage during school hours, answering questions incorrectly once in a while, and quickly answering problems the first time they see the problem, perhaps suggesting that students from suburban schools found the material too easy.

Engaged concentration is largely seen in students from urban middle schools when they pause and take their time at an item followed by answering it correctly, or when they answer a problem on their first attempt rather than requesting hints or scaffolding. The model for students from suburban middle schools mostly detects students to be in engaged concentration when they infrequently request help, or when they answer items slowly but correctly.

The confusion model for students from urban schools largely detects them to be confused when they get successive incorrect answers on a single problem, or when they have incorrectly answered a problem a lot of times in the past and still take a long time to answer it on their next attempt. For the model from students in suburban middle schools, confusion is largely seen in students who frequently request scaffolding on their first attempt, students who use many hints, especially bottom-out hints (final hints that given the answer), and students who make more errors on items involving algebra and fill-in-the-blank types of problems.

Frustration is largely detected in students from urban middle schools when they have repeatedly committed errors on an item and still answer it incorrectly, or when they request hints but still answer incorrectly. For students from suburban middle schools, frustration is mostly detected in students who immediately request a scaffold or hint when answering a question, and in students who take a long time to answer a question.

Off-task behavior detection for middle school students is largely based on students taking long amounts of time between answers, and making relatively few responses. Gaming the system is predominantly seen in students who frequently use large numbers of hints (especially bottom-out hints) and repeatedly access scaffolding when answering a problem.

Student Knowledge

Student knowledge measures were derived from tutor usage in ASSISTments by applying Corbett and Anderson's (Corbett & Anderson, 1995) Bayesian Knowledge Tracing (BKT) model to the interaction or log data (Figure 6). BKT is a knowledge-estimation model which is used in many online learning systems. BKT has been shown in several studies to achieve predictive performance (in terms of predicting future student performance) comparable to or better than competing methods used in online learning (Gong, Beck, & Heffernan, 2010; Pavlik, Cen, & Koedinger, 2009).

BKT is a Hidden Markov Model that aims to infer latent constructs in learning (i.e. does a student know a certain skill at a given time?) from a student's pattern of correct and incorrect answers to problems or problem steps that involve a specific skill (Corbett & Anderson, 1995). Typically, a student who does not know a skill usually gives an incorrect answer when tested on that skill. A student who does know the skill usually gives a correct response. There is, however, a probability that the student will give a correct answer despite not knowing a skill (guess parameter). There is also a possibility that the student will give an incorrect answer despite knowing the skill (slip parameter).

BKT is similar to cognitive diagnosis models (CDMs) in that it makes inferences on the cognitive state (i.e. knowledge). Additionally, one of the most commonly used CDMs, the DINA (deterministic input, noisy, and gate) model, uses guess and slip parameters in estimating the probability a student answers an item correctly. However, there are also significant differences between BKT and CDMs. While CDMs are latent class models that are useful for inferences about cognitive states or processes (Junker & Sijtsma, 2001; Rupp, 2007), CDMs represent compensatory or conjunctive combination of multiple skills per item as latent classes of mastery or non-mastery of those skill patterns. Also, CDMs as a latent factor model does not consider the order in which students solve problems, ignoring the likelihood that performance improves with practice. On the other hand, BKT explicitly incorporates temporal information into its estimates.

In the case of student interaction with ASSISTments, student knowledge is assessed from each student's attempt to answer a problem. Each time a student attempts a problem or problem step for the first time, BKT calculates (and recalculates on next attempt) the estimates of that student's knowledge for the skill involved in that problem or problem step, using four parameters: (1) L_0, the initial probability that the student knows the skill, (2) T, the probability of learning the skill at each opportunity to use that skill, (3) G, the probability that the student will give the correct answer despite showing evidence of not knowing the skill, and (S) the probability that the student will give an incorrect answer despite showing evidence of knowing the skill. The estimates obtained via BKT were calculated at the student's first response to each problem, and were applied to each of the student's subsequent attempts on that problem. In fitting this model to the interaction data, the standard method of using brute-force grid search was used (see Baker et al., 2010).

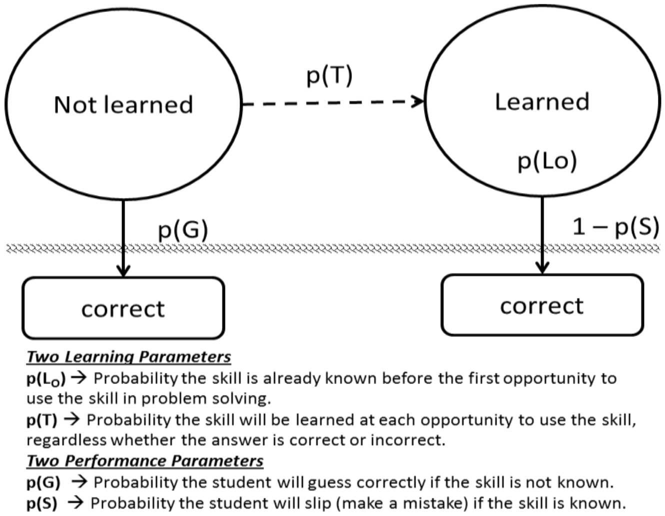


Figure 6. Bayesian Knowledge Tracing (BKT).

Using Bayesian analysis, BKT re-calculates the probability that the student knew the skill before the response (n-1), using the information from the response (help requests are treated as evidence that the student does not know the skill), using these two equations:

$$P(L_{n-1} \mid Correct_n) = \frac{P(L_{n-1}) * (1 - P(S))}{P(L_{n-1}) * (1 - P(S)) + (1 - P(L_{n-1})) * P(G)} \qquad (1)$$

$$P(L_{n-1} \mid Incorrect_n) = \frac{P(L_{n-1}) * P(S)}{P(L_{n-1}) * P(S) + (1 - P(L_{n-1})) * (1 - P(G))} \qquad (2)$$

Then, the system accounts for the possibility that the student learned the skill during the problem step, such that:

$$P(L_n \mid Action_n) = P(L_{n-1} \mid Action_n) + ((1 - P(L_{n-1} \mid Action_n)) * P(T)) \qquad (3)$$

Carelessness

While gaming the system and off-task behavior detectors were trained using data from field observations, carelessness was assessed with a model that detected “slips”, answering a problem incorrectly despite actually knowing how to answer it correctly (Baker, Corbett, & Aleven, 2008; San Pedro, Baker, & Rodrigo, 2011). This is also the same conceptualization as the slip parameter in BKT modeling. Hence modeling carelessness or slip in the context of educational software is derived from BKT where the "contextual slip" model from (Baker, Corbett, & Aleven, 2008; San Pedro, Baker, & Rodrigo, 2011) is used as an operationalization of carelessness. This model infers whether student errors are due to not knowing the skill or due to being careless, based on a combination of the probability of student knowledge (from BKT, discussed above), the pattern of correct and incorrect responses, and other information about the student action (e.g. help-seeking history). We assess contextually the probability of carelessness/slip depending on the context and behavior surrounding the student error. As such, the probability estimate of carelessness/slip is different for each student action.

To model carelessness, BKT is applied to the data to generate initial estimations of whether the student knew the skill at each problem step. Bayesian equations are then used with these estimations (L_n from BKT) to compute the probability of incorrect actions to be slips, based on the correctness or student performance on succeeding attempts to use the skill.

$$P(A_n \text{ } is\ a\ slip \mid A_n \text{ } is\ incorrect) = P(L_n \mid A_{n+1,n+2}) \quad (4)$$

$$P(L_n \mid A_{n+1,n+2}) = \frac{P(A_{n+1,n+2} \mid L_n) * P(L_n)}{P(A_{n+1,n+2})} \quad (5)$$

These probability values are then used to create a model that can predict slip or carelessness contextually at each practice opportunity, from data such as response time, past history, and the pattern and type of errors, without any future information.

CHAPTER IV.

PRELIMINARY WORK

For this dissertation, the author used action-level features from the student sample's interaction data, and utilized the aforementioned models for ASSISTments to obtain measurements of the middle school constructs of interest – cognitive and non-cognitive variables – necessary to address the research questions. The models of student knowledge, academic emotions and student behaviors discussed in Chapter Three were applied to interaction or log data from ASSISTments obtained for a sample of 7,636 middle school students (from both urban and suburban middle schools) who used the system between 2004-2005 and 2008-2009 (Figure 7). The result was a sequence of predictions of student affect and behavior, and estimates of student knowledge across the history of each student's use of the ASSISTment system. This student sample and its data are described in more detail in Chapter Five.

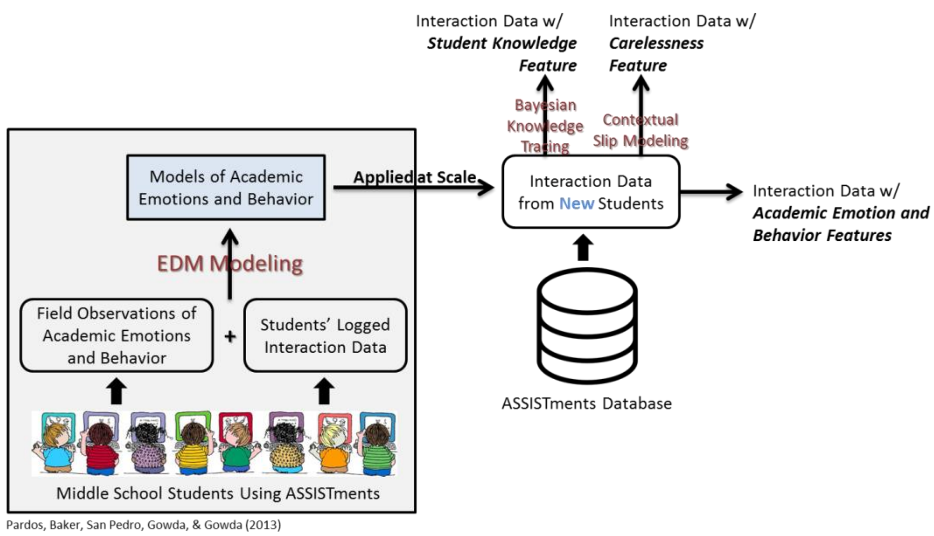


Figure 7. Feature generation in ASSISTments interaction data.

Having obtained fine-grained measures of student knowledge, academic emotions and student behavior during middle school, preliminary studies were conducted for student subsets in the dissertation study's sample (based on not having all the data yet at the time of these studies –

some students in the sample had not graduated high school yet) to analyze the relationships between middle school variables of student behavior, academic emotions and knowledge and individual long-term outcomes (i.e. college attendance or college outcomes). In particular, the author analyzed whether the middle school variables of student behavior, academic emotions and knowledge during a student's middle school learning in ASSISTments were predictive of their eventual college attendance (San Pedro, Baker, Bowers, & Heffernan, 2013). The author also analyzed whether the middle school variables of student behavior, academic emotions and knowledge were predictive of eventual enrollment in a STEM or Non-STEM college major (San Pedro, Ocumpaugh, Baker, & Heffernan, 2014), and their associations to specific college major groups (San Pedro, Baker, Heffernan, & Ocumpaugh, 2015). Lastly, the author also tested the relation of middle school variables of student behavior, academic emotions and knowledge to the eventual enrollment in a selective or not selective postsecondary institution (San Pedro et al., in preparation). This chapter details each of these preliminary studies.

Predicting College Enrollment

In (San Pedro, Baker, Bowers, & Heffernan, 2013), a discovery with models approach was used to study how student learning, academic emotions and behavior in middle school (as assessed by fine-grained measures from interaction data) can predict eventual college enrollment. This study was conducted in a dataset of 3,747 students who used ASSISTments from school years 2004-2005 to 2006-2007, who had completed high school and had the opportunity to enroll in college prior to data collection. These students were drawn from three districts who used the ASSISTments system throughout the year. One district was urban with large proportions of students requiring free or reduced-price lunches due to poverty, relatively low scores on state standardized examinations, and large proportions of students learning English as a second language. The other two districts were suburban, serving generally middle-class populations. Models of student knowledge, academic emotions and behavior were applied to the ASSISTments interaction data for these students, creating features that could be used for the final prediction model of college enrollment. Students' enrollment records were obtained from the National Student Clearinghouse, and these records were used to obtain the variable of whether or not the students in the data set enrolled in college or not.

A final logistic regression model (a more parsimonious model than a full model with all middle school variables) was developed from a combination of variables of academic emotions, behavior and student learning in ASSISTments (Table 2). This model achieved a cross-validated A' (as discussed in Chapter Three) of 0.686, i.e. the model could distinguish a student who will enroll in college from a student who will not enroll in college 68.6% of the time (Fogarty, Baker, & Hudson, 2005; Hanley & MacNeil, 1982). This model was statistically significantly better than the null model, $\chi^2(df = 6, N = 3747) = 386.502$, $p < 0.001$ and had a fit of R^2 (Cox & Snell) = 0.098, R^2 (Nagelkerke) = 0.132, indicating that the model explained 9.8% to 13.2% of

Table 2

Final Model of College Enrollment

Middle School Variables	*Coefficient*	χ^2	*p-value*	*Odds Ratio*
Student Knowledge	1.119	17.696	<0.001	3.062
Correctness	0.698	47.352	<0.001	2.010
Number of First Actions	0.261	28.740	<0.001	1.298
Carelessness	-1.145	28.712	<0.001	0.318
Confusion	0.217	24.803	<0.001	1.242
Boredom	0.169	12.249	<0.001	1.184
Constant	0.351	100.011	<0.001	1.420

the variance in college attendance. For the models, the R^2 values serve as measures of effect sizes; when converted to correlations, they represent moderate effect sizes in the 0.31-0.36 range.

In this model, student knowledge, correctness, number of first actions, boredom, confusion, and carelessness significantly contribute to the overall model of college enrollment. Success within middle school mathematics (indicated by correct answers and high probability of knowledge in ASSISTments) is positively associated with college enrollment, a finding that aligns with studies that find high performance to be a sign of college readiness (Roderick, Nagaoka, & Coca, 2009) and models that suggest student aptitude is predictive of college attendance (Eccles, Vida, & Barber, 2004). For carelessness, once student knowledge in the model is controlled, it becomes negatively associated with college attendance. In other words, once student knowledge is controlled for, careless students are successful but not as successful as they would be expected to be if they weren't careless (cf. Clements, 1982). Also in this model, the likelihood of college enrollment increases with boredom, once the other variables are taken into account (e.g. once student knowledge, tutor usage, and other forms of disengagement are controlled). This may be because after controlling for unsuccessful bored students, all that may remain are students who become bored because the material is too easy (cf. Pardos, Baker, San Pedro, Gowda, & Gowda, 2013). Similarly, after controlling for other variables, confusion is positively associated with college attendance – where after controlling for students who are both confused and unsuccessful, all that is likely to remain may be students who addressed their confusion productively (cf. Lee, Rodrigo, Baker, Sugay, & Coronel, 2011).

Predicting STEM Major Enrollment

Another long-term outcome modeled using the discovery with models approach was whether a student enrolls in a STEM (Science, Technology, Engineering, and Mathematics) major or a Non-STEM major (San Pedro, Ocumpaugh, Baker, & Heffernan, 2014). This model was again created based on middle school variables within the ASSISTments system, and using college major survey data from college students who previously used ASSISTments when they were in middle school. A prediction model was developed to distinguish whether a student enrolled in a STEM major in college or a non-STEM major in college, using assessments of knowledge, academic emotions and engagement from their interaction with ASSISTments.

A total of 425 participants, who had previously used the ASSISTments system for middle school mathematics for one or more years between 2004 and 2007, answered a survey about their post-high school academic and career achievements. Out of the 425 respondents, 363 respondents were in college (85.41%) and they identified the college major they were enrolled in. Interaction data from ASSISTments were obtained for these 363 respondents. Models of student knowledge, academic emotions and behavior were applied to this dataset to be able to develop features used for the final predictive model of STEM major enrollment. This dataset was then labeled to reflect enrollment in a STEM major or not, based on their survey answers of college majors. In this study, STEM majors consist of medical training programs and science and engineering degree programs as defined by the National Science Foundation (National Science Foundation NCSES, 2013).

The final reduced logistic regression (Table 3) achieved a cross-validated A’ of 0.663; the model could distinguish between a student who took a STEM college major and a student who took a non-STEM college major 66.3% of the time. This final model is also statistically significantly better than the null model, χ^2 (df = 2, N = 363) = 38.010, $p < 0.001$, achieving a fit of R^2 (Cox and Snell) = 0.099, R^2 (Nagelkerke) = 0.133.

Table 3

Final Model of STEM Major Enrollment

Features	*Coefficient*	*Chi-Square*	*p-value*	*Odds Ratio*
Student Knowledge	0.357	8.859	0.003	1.429
Gaming	-0.492	13.792	<0.001	0.611
Constant	0.133	1.418	0.234	1.142

This model indicates that the following variables are associated with a lower probability of enrolling in a STEM college major: gaming the system and lower knowledge. Learning within middle school mathematics (indicated by high probability of knowledge in ASSISTments) is positively associated with STEM major enrollment, a finding that aligns with studies that conceptualize high performance and developing aptitude during schooling as a sign of STEM major readiness and enrollment in STEM programs (Wang, 2012; Wang, 2013). The disengaged behavior of gaming the system during middle school mathematics is found to be associated with not pursuing a STEM degree. Previous research has shown that gaming is associated with poorer learning (Cocea, Hershkovitz, & Baker, 2009), but it is also a particularly strong indicator of disengagement with mathematics, suggesting a way that students' lack of interest in STEM careers may manifest early.

Findings in this preliminary study also reveal that academic emotions particularly do not have strong individual effects on whether a student will pursue a STEM or major or not. This is different from the previous preliminary work that found academic emotions to be predictive of college attendance (San Pedro, Baker, Bowers, & Heffernan, 2013). A possible explanation is that academic emotions during schooling largely plays a role in determining whether students choose higher education at all; once only the students who choose higher education are analyzed (e.g. the sample in this preliminary work) academic emotions play a much smaller role than domain-specific learning or choices. This finding does not mean, however, that negative academic emotions during middle school should not be attended to, as they are still associated with both learning outcomes and college attendance (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; San Pedro, Baker, Bowers, & Heffernan, 2013). It may be a valuable area of future work to explore whether the interactions of academic emotions and other factors can influence whether students enroll in a STEM major or a non-STEM major.

Exploring College Major Groups

A related follow-up analysis to (San Pedro, Ocumpaugh, Baker, & Heffernan, 2014) explored how middle school variables were related to what area a student chooses to major in college (San Pedro, Baker, Heffernan, & Ocumpaugh, 2015). Using survey data acquired from 356 college students who used the ASSISTments system when they were in middle school, significant differences in student knowledge, performance, carelessness and gaming behaviors were found between students who eventually choose different college majors. The same data in (San Pedro, Ocumpaugh, Baker, & Heffernan, 2014) that consisted of post-high school survey and interaction data with student learning, academic emotions and behavior features were used in this study. There was a wide variety of responses of college majors, ranging from "General Studies" to "Culinary Arts" to "Criminal Justice." These majors were grouped into eight general classifications developed by The College Board (2014), and each student in the sample was labeled accordingly. One classification that ended up having too few students – the "Trades and Personal Services" – was excluded from this analysis, leaving 356 students with their college majors belonging to the remaining seven classifications, namely: Arts and Humanities, Business, Health and Medicine, Interdisciplinary Studies, Public and Social Services, Science Math and Technology, and Social Sciences.

Overall, findings in this study showed that success within middle school mathematics (i.e., in ASSISTments in this study) is more common in students who eventually enroll in Science, Math and Technology majors than in Business, Interdisciplinary Studies, Public and Social Services, or Social Sciences majors, a finding that aligns with studies that conceptualize high performance and developing aptitude in STEM during schooling as a sign of STEM major readiness and preparation for enrollment in STEM programs (Wang, 2013). In addition, the disengaged behavior of gaming the system during middle school mathematics is found to be associated more with students enrolled in Business, Interdisciplinary Studies, Public and Social Services, and Social Sciences, and less with students enrolled in Science, Math and Technology. The best course of intervention may depend on better understanding this relationship. If gaming reduces the likelihood of pursuing Science, Math and Technology major because it reduces learning, knowledge remediation may be provided – either through alternate opportunities to learn the material that gaming allowed them to bypass or through metacognitive interventions showing why gaming is ineffective for learning (Baker et al., 2006; Arroyo et al., 2007). If gaming is instead an early indicator of lack of interest in STEM, remediation may be more difficult, but the information could still be used to provide actionable reports to teachers about students' likely career interests.

These relationships between middle school student behavior and learning and eventual college major choice show potential in complementing current understanding of why students choose to enroll in different majors. Holland's theory of career choice (1997) asserts that students make academic and career choices compatible with their personality and driven by their preferred activities, interests and competencies. Students who enroll in Science, Math and Technology majors are known to generally prefer practical and concrete (realistic) activities that involve knowledge acquisition and problem solving (investigative) (Holland, 1997; Pike, 2006; Porter & Umbach, 2006) – supporting the finding of success within ASSISTments being related to Science, Math and Technology majors.

Predicting Enrollment in a Selective College

The author also tested and modeled whether students who used ASSISTments when they were in middle school will attend a selective college (San Pedro et al., in preparation). The model was trained on data from 2,732 students who attended college, combining these students' interaction data when they used the system between 2004 and 2008, with data on the selectivity classification of their college institution, taken from Barron's index of college selectivity (College Division of Barron's Education Series, 2012). This resulted in a logistic regression model that could distinguish between a student who will attend a selective college and a student who will not attend a selective college 77.4% of the time (A' = 0.774), with Kappa value = 0.419, when applied to data from new students (Table 4). This model (χ^2(df = 7, N = 2732) = 680.752, $p < 0.001$) indicated that the following middle school variables are associated with a lower probability of attending a selective college: gaming the system, confusion, frustration, gaming the system, less engaged concentration, less carelessness, lower performance and less usage of the system.

Table 4

Final Model of Going to a Selective College

Features	***Coefficient***	***Standard Error***	***Chi-Square***	***p-value***	Odds Ratio
Engaged Concentration	0.119	0.060	3.956	0.047	1.127
Confusion	-0.153	0.064	5.710	0.017	0.858
Frustration	-0.206	0.053	14.907	<.001	0.814
Gaming	-0.186	0.077	5.862	0.015	0.830
Carelessness	0.275	0.081	11.628	0.001	1.316
Correctness	0.835	0.098	72.805	<.001	2.305
Number of Actions	0.200	0.064	9.870	0.002	1.222
Constant	0.404	0.046	76.681	<.001	1.497

The positive connection between academic performance (i.e. correctness) and attending a selective college is consistent with past research using other indicators of academic performance (cf. Baron & Norman, 1992; Carnevale & Rose, 2003; Griffith & Rothstein, 2009). This finding is related to studies that identify college readiness to be linked to high performance during schooling (Roderick, Nagaoka, & Coca, 2009), as well as studies that predict college enrollment to be correlated with indicators of aptitude (Christensen, Melder, & Weisbrod, 1975; Eccles, Vida, & Barber, 2004). Engaged concentration is also indicative of success in attending a selective college, possibly because engaged concentration is widely found to be related to effective learning (Craig, Graesser, Sullins, & Gholson, 2004; D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; Rodrigo et al., 2009). The positive association between attending a selective college and students showing careless behavior, while non-intuitive, may be attributed to careless students who perform well but not as well as they would be if they weren't careless (cf. Clements, 1982). While confusion can sometimes result in successful learning, when confusion is not addressed it is known to be associated with poorer learning (D'Mello & Graesser, 2012). Students who experience frustration and remain in that affective state are less likely to learn (Liu, Pataranutaporn, Ocumpaugh, & Baker, 2013), and can even become bored (D'Mello & Graesser, 2012). It is perhaps not surprising that gaming the system was higher for students who did not attend a selective college, since gaming the system is known to be associated with poorer learning (Cocea, Hershkovitz, & Baker, 2009), poorer performance on standardized state exams (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013), and a lower chance of attending college (San Pedro, Baker, Bowers, & Heffernan, 2013).

Results in each of the preliminary studies support known indicators of successful entry to postsecondary education (academic achievement, grades). They also form a basis for this dissertation. This dissertation utilized the whole and aggregate data spanning across the middle school, high school and college phases. The author made use of and developed the findings from these studies, treating them as preliminary tests for the models this dissertation study aims to create.

CHAPTER V.

METHODOLOGY

The data collection, data processing, and the individual models created and described in Chapter Four form the preliminary findings for this dissertation. The individual models of college attendance, enrollment in a selective college and STEM major enrollment shed light on the student academic emotions and behaviors that occur during the use of learning technology – which are more frequent and in many ways more actionable than the behaviors which result in disciplinary referrals – and how they can be predictive of long-term student outcomes.

This dissertation brought all of this work into a culminating set of models that add to the understanding of the antecedents of college attendance outcomes. To answer the research questions in this dissertation, a longitudinal and more comprehensive investigation of how middle school factors influenced the choices and decisions a student made in entering postsecondary education was conducted with three studies. Figure 8 shows a general view of a model that combines the middle school, high school and college factors. This dissertation presented a cumulative and incremental approach to modeling these factors: Study 1 developed a structural model that aggregated the individual middle school variables of knowledge, academic emotions and behaviors into middle school factors of engagement and performance to predict the college outcomes of college enrollment and selectivity of the college attended, Study 2 modeled the outcome of college major choice from said middle school factors of engagement and performance in Study 1, as well as from a combination of the individual middle school variables. Study 3 explored the potential impact of high school factors on the college outcomes by developing mediation models that included the factors of engagement and performance from middle school, high school factors and college outcomes. Models in each study attempted to address each of the research questions mentioned in Chapter One, and were compared and evaluated for goodness of fit. This chapter includes sections describing the student sample used in this dissertation, the data sources and measures used for the models, and finally the data analyses and modeling conducted for each study.

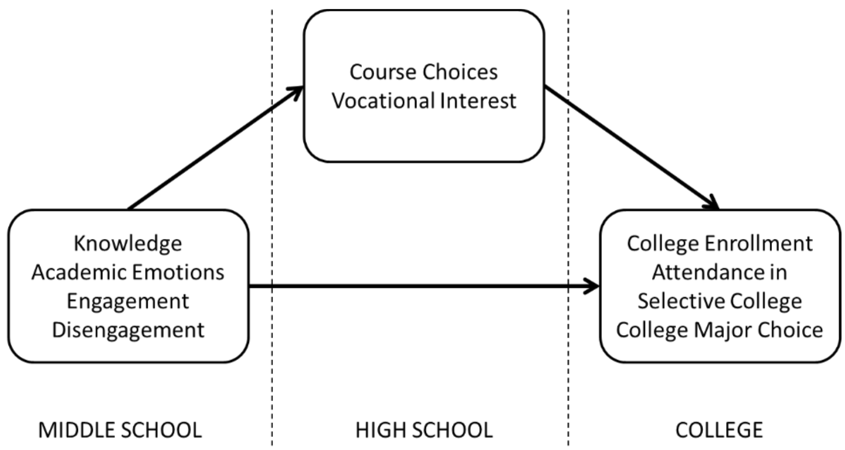


Figure 8. Proposed general model of college attendance outcome.

Student Sample

This dissertation study used the full student sample mentioned in Chapter Four that included 7,636 students who used ASSISTments when they were in middle school from school years 2004-2005 to 2008-2009. These students attended middle school from four districts in the Northeastern United States that used the ASSISTments system throughout the course of a school year (with a small number of students using the tutor for two to three school years).

Two districts were urban with large proportions of students requiring free or reduced-price lunches due to poverty, relatively low scores on state standardized examinations, and large proportions of students learning English as a second language. Both urban districts had below state averages for college readiness (based on percentages of 12th graders who passed in AP/IB exams), and proficiencies for Math and English. Within one urban school district, 40% of the students were Hispanic, 15% were African-America, 7.5% were Asian, and 34% were Caucasian. The second urban district had similar composition except for having fewer African-American students – 47% were Hispanic, 5.4% were African-America, 5.7% were Asian, and 35% were Caucasian.

The other two districts were suburban, serving generally middle-class populations, with relatively higher scores on state standardized examinations. One suburban district had above state averages for college readiness (based on percentages of 12th graders who passed in AP/IB exams), Math and English proficiencies, while the other suburban district had a below state average for college readiness, and average proficiencies for Math and English. Within the first suburban district, 6.3% of the students were Hispanic, 1.9% were African-America, 24% were Asian, and 64% were Caucasian. The second suburban district had similar composition except for Asian students – 5.3% were Hispanic, 3.4% were African-America, 2.1% were Asian, and 88% were Caucasian.

As will be explained in the subsequent sections, the models created for Study 1, Study 2 and Study 3 each derived its modeling sample from this full student sample that used ASSISTments during the middle school years, differing in the outcome variables used for each model (different number of high school and college variables were available for the different student subsets).

Data for College Attendance Pathway Models

To create the pathway models for college attendance outcomes, this dissertation leveraged the existing data obtained from student use of an online learning system in middle school, several years ago, and additional data collected for the same students during high school and after graduation from high school. Student data collected from these sources were integrated.

During middle school, these students used ASSISTments, a tutoring system for middle school mathematics, producing interaction data used to generate features of interest from students' middle school mathematics experiences. Later, data on their course choices and interests in high school, whether they attended college, what college they attended, and what college major they selected were collected.

Students' interactions with ASSISTments were extracted in deidentified form from the ASSISTments database. Then all survey data and college information from the same students were collected and provided to the author of this dissertation in deidentified form. Data linkage was conducted by authorized members of the ASSISTments team. Figure 9 shows these data sources. The variables obtained from each of these data sources were combined to form the dataset that was analyzed in this dissertation using a longitudinal and correlational design. Statistical and data mining models and techniques were used to evaluate the associations between these middle school features and data from high school and college.

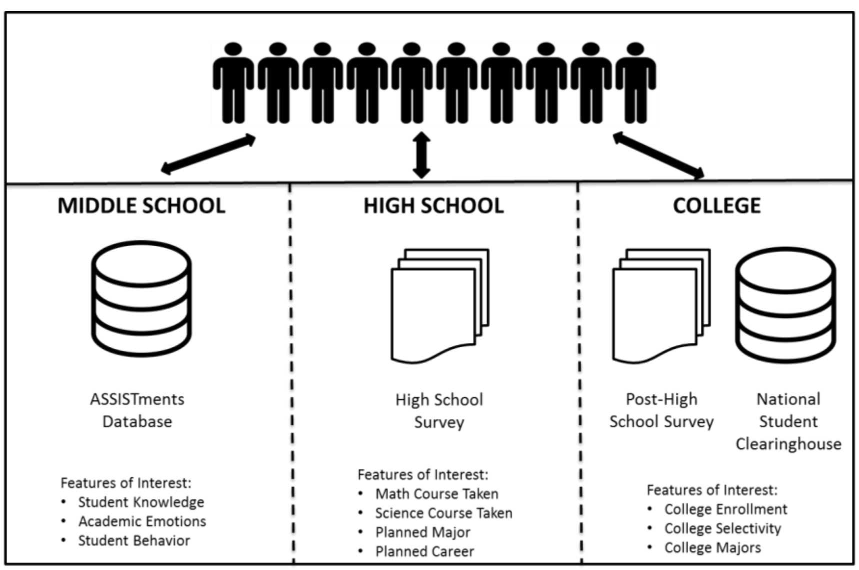


Figure 9. Sources of secondary data.

With the secondary data taken from these data sources, the datasets to be used for this dissertation consisted of middle school, high school, and college variables at the student-level (Table 5). Below are the descriptions of the variables or features extracted from each data source.

Table 5

Student-level Variables from Middle School, High School and College Data

	Middle School Variables	*High School Variables*	*College Variables*
Student	Student Knowledge Academic Emotions: ▪ Boredom ▪ Engaged Concentration ▪ Confusion ▪ Frustration Disengaged Behaviors: ▪ Off-task behavior ▪ Gaming the system ▪ Carelessness Tutor Usage: • Correctness • Number of Actions	AP Math (1 = AP/Honors, 0 = Regular) AP Science (1 = AP/Honors, 0 = Regular) Planned STEM Major (1 = STEM major, 0 = Non-STEM major) Planned STEM Career (1 = STEM career, 0 = Non-STEM career)	College Enrollment (1 = Enrolled, 0 = Not enrolled) Selectivity of college attended (Ordinal) College Major Choice (1 = STEM major, 0 = Non-STEM major)

Middle School Data

The ASSISTment system was the primary source of middle school data. As mentioned in Chapter Four, the middle school variables were derived by applying models of student knowledge, academic emotions and student behaviors to interaction or log data obtained from the sample of 7,636 middle school students who used the ASSISTment system between 2004-2005 and 2008-2009. Overall, these students made over 6 million actions within the software (where an action consisted of making an answer or requesting help), within an estimated total of over 2 million mathematics problems (counting both original and scaffolding problems), working on an average of over 250 problems per student.

The middle school variables consisted of academic emotions, behaviors, knowledge, and student usage aggregated (i.e. averaged) at the student-level. The academic emotions consisted of four variables: boredom, engaged concentration, confusion, and frustration. The student behaviors consisted of three variables: off-task behavior, gaming the system, and carelessness. As explained in previous chapters, the values for these variables, together with student knowledge, were derived from model prediction confidence values, and thus, were continuous/numeric. Each of these middle school variables of performance and engagement were inferred using the interaction-based models discussed in Chapter Three. For example, confusion can be inferred from students from urban schools who use ASSISTments, when their encounter successive incorrect answers on a single problem, or when they have incorrectly answered a problem a lot of times in the past and still take a long time to answer it on the next attempt. Student usage consisted of percentage of correctness (whether responses were correct or incorrect) and number of actions made by the students. These actions were also derived from students' interaction data, and were also numeric (i.e. count data for number of actions).

High School Data

Students who used ASSISTments during their middle school years and who were in high school at the time of data collection were administered a survey. The survey was a short questionnaire that asked the highest level of math and science courses that the student completed in high school and asks the student what his/her educational and career plans are upon graduation. Students who used ASSISTments when they were in middle school during school years 2007-2008 and 2008-2009 completed the questionnaire between the fall of 2012 and the spring of 2013. Around 282 students were identified to be attending high schools in the same urban or suburban districts in New England. Surveys were not given to all high schools in the larger urban districts, just the high schools with the largest proportion of these students. They were given short paper questionnaires on high school course taking during their regular classroom time and college major and career of interest (Appendix B shows the high school survey questions). This survey asked the following questions:

- Most recent high school math course taken
- Most recent high school science course taken
- Current or past employment while in high school
- Postsecondary plans (work or college)
- College major of interest if planning to go to college
- Career of interest after high school

They were then invited to answer and complete an optional online survey (CAPAExplore survey) when they went home, which further explores their interest and confidence in courses. Relatively few students completed this additional survey, and it is outside the scope of this dissertation.

Four high school variables from the high school survey were used for this dissertation – AP Math, AP Science, Planned STEM major, and Planned STEM career, all of which were coded in binary format. AP Math values were based on the question, “What mathematics course are you taking right now? If you are not taking a mathematics course now, what is the last (most recent) mathematics course you took?” The answers to this survey question varied from regular type of math course (ex. Discrete Math, Math 4) to AP or Honors type of math course (ex. AP Statistics, Honors Calculus) When the student’s most recent math course was regular type, AP Math was coded as 0. When the student’s most recent math course was either AP Math or Honors Math, AP Math was coded as 1.

AP Science values were based on the question, “What science course are you taking right now? If you are not taking a science course now, what is the last (most recent) science course you took?” The answers to this survey question varied from regular type of science course (ex. Environmental Science, Physics) to AP or Honors type of science course (ex. AP Biology, AP Chemistry) When the student’s most recent science course was regular type, AP Science was coded as 0. When the student’s most recent science course was either AP Science or Honors Science, AP Science was coded as 1.

Planned STEM major values were based on the survey question, “If you plan to go to college after high school, what major or majors do you find most interesting?” The answers were coded to STEM major (value of 1) or non-STEM major (value of 0), defined as medical training programs and degree programs eligible for National Science Foundation STEM funding (National Science Foundation NCSES, 2013).

Planned STEM career values were based on the survey question, “If you plan to work after high school (whether or not you go to college), what kind of jobs are you interested in?” The answers were coded to STEM career (value of 1) or non-STEM career (value of 0).

The high school course choice variables (AP Math, AP Science) were limited by the free response in the existing survey data about the students’ most recent or current math or science class. This did not take into account that a student may have taken AP course before the current class, i.e. they take AP Physics as a junior, take Regular Biology as a senior. Also, the student respondents of the high school surveys belonged to high schools that offered AP or Honors Math and Science courses. However, these high schools differ in the types of AP or Honors Math and Science courses that they offer, a further limitation of the existing high school survey data gathered.

College Data

Data was collected on the postsecondary education status of this dissertation's student sample expected to be in the postsecondary stage of education by the time of data collection. For their college enrollment information, records were requested from the National Student Clearinghouse. This information was supplemented with college selectivity classification of the postsecondary institutions, taken from the Barron's College Selectivity Rating which classifies colleges into ten categories, from most selective or 'Most Competitive' to 'Non-Competitive' to 'Special'. Another source of data during this phase included survey data about post-high school academic and career placement that was administered to a subset of students. Each type of postsecondary data here was similar to the data used in the preliminary studies in Chapter Four. This dissertation aggregated these data as a whole to form the college attendance outcomes used in Studies 1 to 3. Three outcome variables were derived from these data sources: college enrollment (dichotomous, 0 = not enrolled in college, 1 = enrolled in college), selectivity of college attended (ordinal, 0 = Unclassified to 10 = Most Selective), and college major type (dichotomous, 0 = Non-STEM major type, 1 = STEM major type).

Postsecondary Institution Data. College enrollment records for 2013 for the student sample (7,636 students) were obtained from the National Student Clearinghouse (NSC, http://www.studentclearinghouse.org). This data included whether a student was enrolled in a college or not, the name of the university, date of enrollment, and college major enrolled in if available (however, this information was seldom available). As mentioned, the author of this dissertation received deidentified postsecondary data from the ASSISTments team. As discussed in Chapter Four, a subset of the student sample was used in a preliminary study to predict college enrollment from interaction with ASSISTments. Compared to that previous study, this dissertation now included students who used ASSISTments in subsequent years (2007-2008 to 2008-2009). The data from NSC included all records of a student's enrollment in post-secondary institutions, accounting for students who transferred colleges. Only the most recent post-secondary institution the student enrolled in was used in the dataset for this dissertation, to provide one outcome of postsecondary institution for each student. Additional information (such as whether the student graduated from college) is generally available from the Clearinghouse, but was not available for these students for a few more years.

Selectivity Measure. Selectivity measures of post-secondary institutions are generally determined by an aggregate computed across several factors, including: the median SAT or median composite ACT entrance exam score; the average high school class rank of the student; the average student GPA in high school; and the percentage of students accepted (Carnevale and Rose, 2003). The most commonly-used measure of college selectivity (c.f., Carnevale & Rose, 2003; Schmidt, Burroughs, Cogna, & Houang, 2011) is the Barron's index (College Division of Barron's Education Series, 2012). They also offer a publicly available longitudinal database (http://www.barronspac.com) containing this information. The Barron's College Selectivity Rating classifies colleges into ten categories (Schmidt, Burroughs, Cogna, & Houang, 2011; College Division of Barron's Education Series, 2012) from most selective or 'Most Competitive' to 'Non-Competitive' to 'Special'. Colleges under 'Special', consist of specialty institutions such as schools of music, culinary schools, automotive training schools, and art schools (ex. New England Conservatory of Music). This is summarized in Table 6 below. 'Special' institutions represent institutions that select students based on fundamentally different criteria than the other institutions studied. Hence, for this study these 'Special' institutions were labeled without any level of selectivity – the same label as those that were unclassified.

Table 6

Barron's College Selectivity Rating

Selectivity Rating	*Selectivity Description*	*Example Institution(s)*	*Number of Students*
10	Most Competitive	Columbia, Harvard, Stanford	188
9	Highly Competitive+	Cornell University	152
8	Highly Competitive	Fordham University	152
7	Very Competitive+	Yeshiva University	40
6	Very Competitive	Hunter College	474
5	Competitive+	Buffalo State College	50
4	Competitive	St. Joseph's College	1154
3	Less Competitive	Berkeley College	136
2	Non-Competitive	College of Staten Island	1653
1	Special	Julliard School	42
(1)	(Unclassified)	Glendale Community College	90

Post-High School Survey Data. Over 2,500 students who had used ASSISTments during their middle school mathematics classes were identified and invited to participate in a survey (Appendix C) about their post-high school academic and career achievements. A total of 425 students responded, for a retention rate of about 20%. This proportion was obtained through considerable effort. The first step, which was relatively unsuccessful, was to advertise the online survey through social media venues. After several months, the last known addresses from the school districts were obtained and paper surveys were sent (which also included instructions that allowed students to answer online). Finally, a consultant from the cohort of students was hired to help reach non-respondents. Of the respondents, 62% responded through an online survey and 38% through U.S. mail. These students were drawn from three school districts in the Northeastern United States and used ASSISTments during the 2004-2005 to 2006-2007 school years (with a few continuing tutor usage for more than one year). Within the survey, students were asked to specify what degree program(s) they were enrolled in, whether they were engaged in full or part-time employment, and what their current employment was.

Of 425 respondents, 363 indicated that they were enrolled in a degree program. A wide variety of responses was received, ranging from "General Studies" to "Culinary Arts" to "Criminal Justice." These majors were grouped into eight general classifications provided by The College Board (2014) (Table 7), and each student in the sample was labeled accordingly. These college major groups include: Arts and Humanities, Business, Health and Medicine, Interdisciplinary Studies, Public and Social Services, Science Math and Technology, Social Sciences, and Trades and Personal Services. Each of these college majors was classified whether it was a STEM or non-STEM major, defined as medical training programs and degree programs eligible for National Science Foundation STEM funding (National Science Foundation NCSES, 2013).

Table 7

College Major Classifications by the College Board

Major Categories	*Number of Students*
Arts and Humanities	26
Business	56
Health and Medicine	52
Interdisciplinary Studies	24
Public and Social Services	23
Science, Math and Technology	92
Social Sciences	82
Trades and Personal Services	8

Available measures from middle school, high school and college were combined into a single integrated data set for the 7,636 student sample (Figure 10). The integrated data set showed that the entire 7,636 students had measures for middle school variables and college outcomes of college enrollment and selectivity of college attended. Out of the 7,636 students, a smaller subset of 363 students who enrolled in college also had the college outcome of college major from the post-high school survey. Different from that subset was another smaller subset of 282 students that had data for all the phases – having information for all middle school variables, all high school variables, and two college outcomes of college enrollment and selectivity of college attended. With the large missing data for the high school variables and college major outcome (95% or more of entire student sample), data imputation may be less reliable given the small number of students with observed high school variables and observed college major outcome. Hence, these three student datasets formed the basis of the analyses and modeling conducted in Studies 1 to 3.

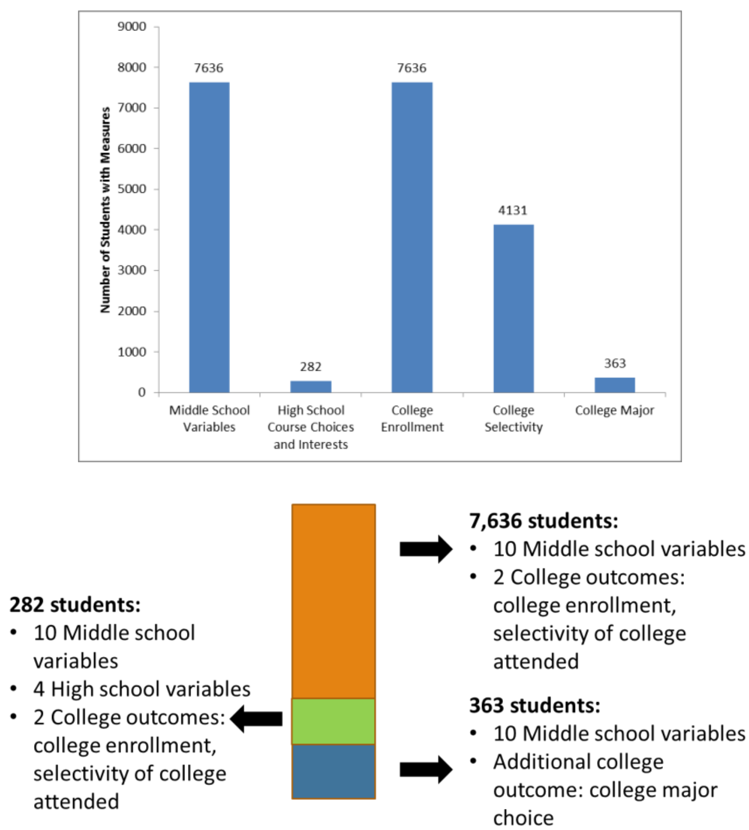


Figure 10. Number of students with middle school, high school and college variables.

Modeling College Attendance Outcomes

Data assumptions in creating structural models (i.e. SEM) were tested with the measures or variables for this dissertation. With the data imbalance previously mentioned, three studies were conducted to create models with varying outcomes and student sample sizes – Study 1 conducted SEM analysis using student data that have information for only the middle school and college years (n = 7,636 students), Study 2 modeled the college outcome of major choice from middle school information (n = 363 students), and Study 3 created mediation models that used information from middle school, to high school, to college years (n = 282 students). There is no overlap in the students used in Study 2 and Study 3. The resulting models were evaluated and interpreted through their goodness-of-fit measures and parameter estimates. *MPlus 7* software was used in creating the models for this dissertation.

Study 1: Middle School → College Model (College Enrollment and Selectivity of College Attended)

The research question for this study – *Are student behavior, academic emotions and knowledge during middle school computer-based math learning predictive of college enrollment and selectivity of the college attended?* – was answered by modeling whether a student enrolled in college or not, and the selectivity of the college the student attended using the student's interaction-based measures of knowledge, academic emotions and behavior when the student was in middle school and used the ASSISTments system.

Procedures for structural equation modeling (SEM) were used for Study 1 as is often used in research that models factors contributing to constructs relevant in SCCT. For example, Nugent and colleagues (2015) used SEM analyses to model how instructional, motivational and social factors contribute to STEM learning and career orientation, using SCCT as a conceptual framework. Wang (2013) uses SEM to draw upon SCCT and higher education research model entrance into STEM majors by recent high school graduates attending college. Luse and colleagues (2014) utilized SEM to demonstrate how both interest and outcome expectations are positive associated with choice on major. Hence, in Study 1, structural models were tested and evaluated to address this first research question, using the middle school variables as the independent variables (Table 8) from which middle school factors of performance and engagement were derived (explained below), and college enrollment and college selectivity as dependent variables (Table 9) for 7,636, students.

From Table 8, the middle school variables of knowledge, correctness, boredom, engaged concentration, confusion, frustration, off-task, gaming and carelessness have values ranging from 0 to 1, with mean values over these nine constructs of as low as 0.079 (confusion) and as high as 0.647 (engaged concentration). The middle school variable number of actions have original count values from 2 to 14,378. Each of the middle school variables followed a non-normal distribution, most of them being positively skewed. Log transformations of each variable still resulted in non-normality (according to Kolmogorov-Smirnov tests). Thus, in Study 1, I used a model estimator that is robust to the original non-normal values – with the exception of the variable for number of actions. Initial models created resulted in non-convergence when I used the original values of the number of actions variable, given its large kurtosis compared to the rest of the middle school variables. Using the log transform of the variable number of actions resulted in convergence and this transformation was eventually used throughout Studies 1 to 3. The models created throughout Studies 1 to 3 were fit using maximum likelihood estimation with robust standard errors (MLR estimator in *MPlus*). MLR was used for its robustness to non-normality for both continuous and categorical outcomes (Savalei, 2010).

Table 8

Descriptive Characteristics of Middle School and College Variables for Study 1 (n = 7,636 students)

	Mean	*Min*	*Max*	*Std. Dev*	*Skewness*	*Kurtosis*
Knowledge	0.355	0.027	0.954	0.224	0.630	-0.777
Carelessness	0.227	0.008	0.881	0.150	0.998	0.232
Correctness	0.420	0.000	0.964	0.149	0.685	0.017
Number of Actions						
Original	829.538	2	14378	902.370	3.795	26.834
Log Transform	2.732	0.301	4.158	0.419	-0.340	0.650
Boredom	0.214	0.012	0.487	0.075	-0.912	0.159
Engaged Concentration	0.647	0.337	0.949	0.067	-0.033	3.128
Confusion	0.079	0.000	0.605	0.053	0.739	3.132
Frustration	0.174	0.000	0.749	0.110	0.622	0.321
Off-Task	0.210	0.053	0.837	0.085	1.544	4.126
Gaming	0.188	0.001	0.841	0.160	1.078	0.700

For the categorical college outcome variables for Study 1, Table 9 shows that college enrollment had similar frequency and distribution for its two values – enrolled (n = 4131, 54.1%) and not enrolled (n = 3505, 45.9%). For the ordinal variable of selectivity of college attended, those that were either not enrolled in any college, enrolled in a specialized college (e.g. culinary, aviation school) or unclassified by Barron's, had the highest frequency in the sample (n = 3637, 47.6%).

Table 9

Frequency of College Outcomes for Study 1(n = 7,636 students)

	Value	*n*	*%*
Enrollment			
Not Enrolled	0	3505	45.9
Enrolled	1	4131	54.1
Selectivity of College Attended			
Not Enrolled, Special, or Unclassified	0	3637	47.6
Non-Competitive	1	1653	21.6
Less Competitive	2	136	1.8
Competitive	3	1154	15.1
Competitive+	4	50	0.7
Very Competitive	5	474	6.2

Very Competitive+	6	40	0.5
Highly Competitive	7	152	2
Highly Competitive+	8	152	2
Most Competitive	9	188	2.5

Further data analyses were first conducted on the middle school and college variables used in Study 1 that included looking into the correlations between middle school variables, and the relations of middle school variables with respect to the college outcomes (through simple logistic regression).

After looking at the individual middle school variables and their relations to college attendance outcomes, structural equation modeling was conducted to see how middle school factors of engagement and performance (i.e. performance-engagement factors) derived from the ten middle school variables can predict college attendance outcomes of college enrollment and selectivity of college attended. First, a factor structure for the ten middle school variables was identified. This was determined by conducting Principal Component Analysis (PCA) with a Varimax (orthogonal) rotation in *SPSS 19* to identify the underlying structure among the ten middle school variables. A middle school performance-engagement factor (or component, in terms of PCA) was then defined in terms of middle school variables that had rotated loading values of 0.4 and above for that factor.

One reason to use performance-engagement factors in middle school as predictors of college outcomes instead of the ten middle school variables was for dimension reduction (especially in the presence of multicollinearity), and test if the college outcomes can be effectively modeled with a smaller number of predictors without losing much information. Each of these performance-engagement factors may be measured by the combined effects of the individual variables in middle school. Identifying these middle school factors also allowed me to represent underlying concepts or constructs that existed during the student's middle school learning experience with ASSISTments that may be predictive of the college outcomes.

College enrollment and selectivity of college attended were then modeled from the resulting middle school performance-engagement factors using SEM analysis with formative factors (instead of the traditional reflective factors in SEM), where each factor was predicted by respective middle school variables determined in the previous PCA (Figure 11); as such, Study 1 models use the structure out of the previous PCA. Two versions of this model were created for comparison – one where the middle school performance-engagement factors were treated as latent or unobserved and were fit from their corresponding individual middle school variables ("unconstrained"), and another where the middle school performance-engagement factors were not treated as latent or were observed, represented by the component scores from the previous PCA ("constrained").

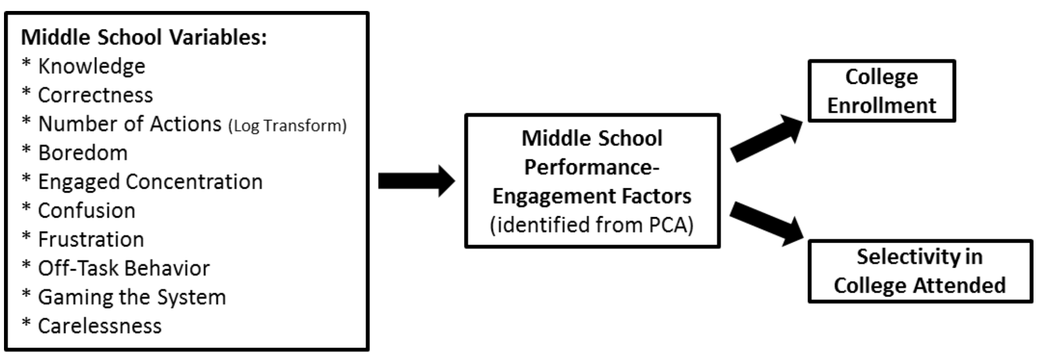


Figure 11. Model design of college enrollment and selectivity of college attended.

Given that Study 1 modeled two college outcomes for the entire student sample, I also created a model for the sole college outcome of attending a selective college for comparison. This model used the 4,131 students from the entire student sample who were enrolled in college.

Study 2: Middle School → College Model (College STEM Major)

Similar to Study 1, the research question for this study – *Are student behavior, academic emotions and knowledge during middle school computer-based math learning predictive of the pursuit or choice of a STEM college major once students are in college?* – was answered by using interaction-based measures of knowledge, academic emotions and behavior from 363 students who used the ASSISTments system when they were in middle school, to model the college outcome of college major choice (STEM or non-STEM major). The ten middle school variables of the 363 students showed similar characteristics as in Study 1 (Table 10). As in Study 1, the log transform values of the variable number of actions were used in the models created (with MLR estimator in *MPlus*).

Table 10

Descriptive Characteristics of Middle School Variables for Study 2 (n = 363 students)

	Mean	Min	Max	Std. Dev	Skewness	Krutosis
Knowledge	0.407	0.061	0.940	0.210	0.382	-0.936
Carelessness	0.243	0.035	0.799	0.137	0.949	0.456
Correctness	0.487	0.089	0.857	0.156	0.159	-0.596
Number of Actions						
Original	910.750	72	14378	1223.894	6.851	62.193
Log Transform	2.798	1.857	4.158	0.356	0.203	0.719
Boredom	0.216	0.027	0.336	0.072	-1.234	0.472
Engaged Concentration	0.658	0.456	0.925	0.054	0.960	6.642
Confusion	0.075	0.000	0.195	0.043	-0.124	-0.478
Frustration	0.165	0.006	0.418	0.091	0.333	-0.095
Off-Task	0.207	0.067	0.546	0.071	1.287	2.627
Gaming	0.139	0.004	0.750	0.149	1.658	2.457

Table 11 shows that 194 students from the sample enrolled in a STEM major (53.4% of students who had college major information), while 169 students enrolled in a non-STEM major (46.6%).

Table 11

Frequency of College Major Choice for Study 2 (n = 363 students)

	Value	*n*	%
College Major Choice			
Non-STEM Major	0	169	46.6
STEM Major	1	194	53.4

For Study 2, I conducted logistic regression using the ten middle school variables to determine what combination of interaction-based measures has the best predictive power (Figure 12.a). I also modeled college major choice using the resulting component scores of the middle school performance-engagement factors from Study 1 to see how these factors can also be predictive of this particular college attendance outcome (Figure 12.b).

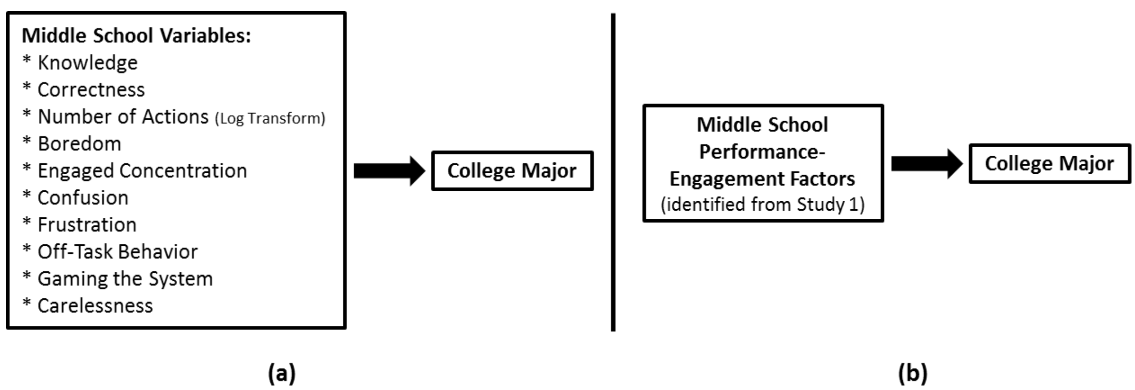


Figure 12. Model design of college major choice (STEM vs. Non-STEM): (a) Using middle school variables only; (b) Using middle school performance-engagement factors from Study 1.

Study 3: Middle School → High School → College Model

Research question 3 – *How do high school course choices and interests in college majors and career during high school mediate between student behavior, academic emotions and knowledge in middle school computer-based math learning, and college attendance outcomes?* – was answered by testing the mediational effects of high school variables AP Math, AP Science, planned STEM major, and planned STEM career on the relationships between the middle school performance-engagement factors and college outcomes of college enrollment and selectivity of college attended for 282 students, giving a broader look at modeling middle school to high school to college. Mediational analysis were employed for Study 3 to evaluate a pathway from middle school factors to high school variables to college outcomes. Such analysis has been successful at modeling cognitive and non-cognitive constructs in education over a period of time (Blackwell, Trzesniewski, & Dweck, 2007). For Study 3, I used the middle school performance-engagement factors instead of the individual middle school variables since they were the predictors of college outcomes used in Study 1. Study 3 used the component scores generated from Study 1 for the 282 students to represent the middle school performance-engagement factors.

Table 12 shows the frequency characteristics of the categorical high school and college variables used for Study 3. Most of the student sample had enrolled in college (n = 225) and in a competitive college (n = 72). When they were in high school, half of the student sample took regular math (n = 141) and science (n = 152) courses, and the other half took AP/Honors math (n = 141) and science (n = 130) courses. From their high school survey information, 156 of the students were interested in taking a non-STEM major in college, while 126 students were interested in taking a STEM major in college (planned STEM major). Despite this, 217 students stated they were not planning to pursue a STEM career, and only 65 students stated they were planning to pursue a STEM career (planned STEM career) – which may be an indication that the students were not certain yet of their job plans after college, at the time of data collection.

Table 12

Frequency of High School and College Variables for Study 3(n = 282 students)

	Value	*n*	*%*
College Enrollment			
Not Enrolled	0	57	20.2
Enrolled	1	225	79.8
Selectivity of College Attended			
Not Enrolled, Special, or Unclassified	0	60	21.3
Non-Competitive	1	44	15.6
Less Competitive	2	6	2.1
Competitive	3	72	25.5
Competitive+	4	11	3.9
Very Competitive	5	45	16.0
Very Competitive+	6	7	2.5
Highly Competitive	7	14	5.0
Highly Competitive+	8	9	3.2
Most Competitive	9	14	5.0
High School Math Course			
Regular	0	141	50.0
AP/Honors	1	141	50.0
High School Science Course			
Regular	0	152	539.
AP/Honors	1	130	46.1
Planned STEM Major			
Non-STEM Major	0	156	55.3
STEM Major	1	126	44.7
Planned STEM career			
Non-STEM Career	0	217	77.0
STEM Career	1	65	23.0

The model evaluated in this study (Figure 13) was designed to correspond to SCCT where learning experiences influence the development of self-efficacy and interests which in turn influence the choices (i.e. college attendance) made by students. Thus, the model developed here aimed to establish if the relationships between college attendance and middle school performance-engagement factors measured within computer-based learning were mediated by high school variables of course-taking (taking regular or AP/Honors math and science courses) and postsecondary interest (plans to pursue a STEM or non-STEM major or career). To evaluate these possible mediations, associations were first established between the middle school performance-engagement factors and outcomes of college enrollment and selectivity in college attended, between middle school performance-engagement factors and high school course choice and postsecondary interest, and between high school course choice and postsecondary interest and outcomes of college enrollment and selectivity in college attended. When significant relationships were found in all these three stages between middle school performance-engagement factors, high school choice and interest, and college outcomes, mediation was tested from middle school to high school to college.

The mediation models created (with MLR estimator in MPlus) used the Sobel method (MacKinnon, 2008) in testing the significance of any mediation found, called the indirect effect. This indirect effect is represented by the product of two coefficients in the model created: the coefficient of middle school performance-engagement factor predicting high school course choice or postsecondary interest, and the coefficient of high school course choice or postsecondary interest predicting the college attendance outcomes. The Sobel method provides an estimate of the standard error of this product and tests if it's different from zero (MacKinnon, 2008).

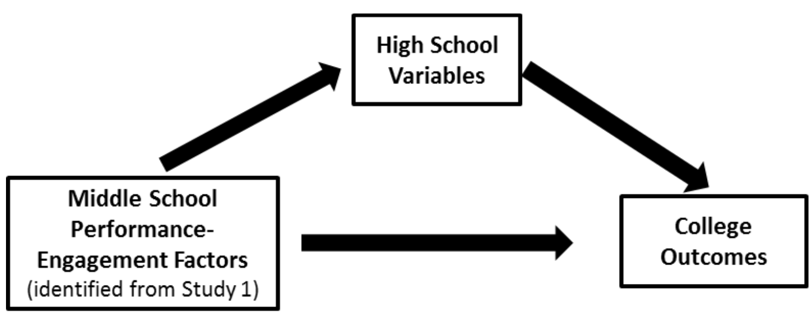


Figure 13. Model design of college attendance outcomes with mediation from high school.

CHAPTER VI.

RESULTS

Study 1: Middle School → College Model (College Enrollment and Selectivity of College Attended)

Correlations between the ten middle school variables were evaluated to establish if there was an existing factor structure. Table 13 shows the strongest positive associations were found between student knowledge and carelessness (r = 0.958, p<.001), student knowledge and correctness (r = 0.783, p<.001), and confusion and boredom (r = 0.693, p<.001). Conversely, the strongest negative associations were between off-task and number of actions (r = -0.637, p<.001), correctness and gaming (r = -0.614, p<.001), off-task and gaming (r = -0.528, p<.001). With these significant correlations between individual middle school variables, a factor structure for these variables was evaluated as basis for the middle school performance-engagement factors used in Study 1 college models.

Table 13

Correlations between Study 1Middle School Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Knowledge	1									
(2) Carelessness	.958**	1								
(3) Correctness	.783**	.661**	1							
(4) Number of Actions (log)	.118**	.214**	-.181**	1						
(5) Boredom	-.372**	-.448**	-.114**	-.518**	1					
(6) Engaged Concentration	.220**	.260**	.080**	.474**	-.515**	1				
(7) Confusion	-.445**	-.455**	-.310**	-.408**	.693**	-.376**	1			
(8) Frustration	-.066**	0.006	-.245**	.116**	.374**	-.152**	.293**	1		
(9) Off-Task Behavior	.089**	.024*	.253**	-.637**	.427**	-.356**	.243**	-.042**	1	
(10) Gaming the System	-.301**	-.196**	-.614**	.506**	-.303**	.323**	-.198**	.169**	-.528**	1

* - p < 0.05; ** - p< 0.001

After examining the correlations between the ten middle school variables, the relationships of college enrollment and selectivity of college attended with each of the middle school variables were evaluated. Simple logistic regression was first conducted on each of the middle school variables to assess their individual relationship with whether a student enrolled in college or not. Table 14 shows that a student was more likely to enroll in college when a student showed more knowledge (β = 0.290, *SE* = 0.025, p<.001, Odds Ratio = 1.336), achieved more correct answers (β = 0.490, *SE* = 0.025, p<.001, Odds Ratio = 1.633), produced more actions in the tutor (β = 0.160, *SE* = 0.023, p<.001, Odds Ratio = 1.173), was more engaged during tutor usage (β = 0.102, *SE* = 0.026, p<.001, Odds Ratio = 1.107), or produced more careless errors (β = 0.194, *SE* = 0.023, p<.001, Odds Ratio = 1.214). On the other hand, a student was less likely to enroll in college, the more a student was bored (β = -0.081, *SE* = 0.023, p<.001, Odds Ratio = 0.922), the more a student was confused (β = -0.167, *SE* = 0.023, p<.001, Odds Ratio = 0.846), the more a student was frustrated (β = -0.274, *SE* = 0.024, p<.001, Odds Ratio = 0.760), or the more a student gamed the system (β = -0.387, *SE* = 0.024, p<.001, Odds Ratio = 0.679). These findings are in line with the preliminary work in Chapter Four that used a smaller student sample (San Pedro, Baker, Bowers, & Heffernan, 2013).

Table 14

Simple Logistic Regression Models of College Enrollment for Each Middle School Variable (Standardized Model Results)

#	*Middle School Variable*	*Estimate*	*S.E.*	*p-value*	*Odds Ratio*
1	Knowledge	**0.290**	**0.025**	**<.001**	**1.336**
	Constant	0.169	0.023	<.001	1.184
2	Correctness	**0.490**	**0.025**	**<.001**	**1.633**
	Constant	0.180	0.024	<.001	1.197
3	Number of Actions (log)	**0.160**	**0.023**	**<.001**	**1.173**
	Constant	0.165	0.023	<.001	1.180
4	Boredom	**-0.081**	**0.023**	**<.001**	**0.922**
	Constant	0.165	0.023	<.001	1.179
5	Engaged Concentration	**0.102**	**0.026**	**<.001**	**1.107**
	Constant	0.165	0.023	<.001	1.179
6	Confusion	**-0.167**	**0.023**	**<.001**	**0.846**
	Constant	0.165	0.023	<.001	1.180
7	Frustration	**-0.274**	**0.024**	**<.001**	**0.760**
	Constant	0.166	0.023	<.001	1.181
8	Off-Task Behavior	0.043	0.023	0.063	1.044
	Constant	0.164	0.023	<.001	1.179
9	Gaming the System	**-0.387**	**0.024**	**<.001**	**0.679**
	Constant	0.165	0.023	<.001	1.180
10	Carelessness	**0.194**	**0.023**	**<.001**	**1.214**
	Constant	0.166	0.023	<.001	1.181

Table 15 shows the correlations of selectivity of college attended with the individual middle school variables. Spearman rank order correlation was used because it determines the correlation between sets of ranked data (ordinal outcome variable of selectivity of college attended). Selectivity of college attended was significantly correlated with all the middle school variables, with positive associations with knowledge (r = 0.270, p<.001), correctness (r = 0.358, p<.001), number of actions (r = 0.092, p<.001), engaged concentration (r = 0.147, p<.001), carelessness (r = 0.209, p<.001) and off-task behavior (r = 0.074, p<.001). Surprisingly, selectivity of college attended showed a weak but significant positive correlation with off-task behavior. Conversely, negative correlations were found between selectivity of college attended and boredom (r = -0.036, *p*<.001), confusion (r = -0.148, *p*<.001), frustration (r = -0.169, *p*<.001) and gaming the system (r = -0.254, *p*<.001).

Table 15

Correlations between Selectivity of College Attended and Middle School Variables

	Knowledge	*Correctness*	*Number of Actions (log)*	*Boredom*	*Engaged Concentration*
Selectivity of College Attended	0.270**	0.358**	0.092**	-0.036**	0.147**
	Confusion	*Frustration*	*Off-Task Behavior*	*Gaming the System*	*Carelessness*
Selectivity of College Attended	-0.148**	-0.169**	0.074**	-0.254**	0.209**

* - p < 0.05; ** - p< 0.001

Another interesting finding in this correlational analysis in Study 1 is the low degree of significance found between off-task behavior and the college attendance outcomes. This is a notable contrast to classroom practice and behavioral support that address off-task behavior. This finding does not mean that off-task behavior during middle school should not be attended to, as it still impacts learning outcomes (Baker, 2007; Cocea, Hershkovitz, & Baker, 2009; Rowe, McQuiggan, Robison, & Lester, 2009).

Identification of Middle School Performance-Engagement Factors

Principal Component Analysis (PCA) with a Varimax (orthogonal) rotation was conducted on the ten middle school variables. The Kaiser-Meyer Olkin measure of sampling adequacy was 0.692 suggesting that the sample for Study 1 is adequate for PCA to produce distinct components or factors. The resulting components and their loadings are shown in the rotated component matrix in Table 16. This shows the estimates of the correlations between each of the middle school variables and the estimated components.

Table 16

Rotated Component Matrix for Principal Component Analysis of Middle School Variables

	Component			
	1	2	3	4
Knowledge	**0.973**	0.021	-0.046	0.077
Carelessness	**0.954**	-0.107	0.005	0.073
Correctness	**0.804**	**0.405**	-0.197	0.111
Number of Actions (log)	0.148	**-0.829**	0.054	0.219
Boredom	-0.366	**0.611**	**0.557**	-0.140
Engaged Concentration	0.170	-0.390	-0.170	**0.859**
Confusion	**-0.497**	**0.503**	**0.536**	0.092
Frustration	0.030	-0.180	**0.928**	-0.138
Off-Task Behavior	0.098	**0.767**	0.018	-0.187
Gaming the System	-0.346	**-0.797**	0.064	0.051

Table 17 shows the total variance explained for each component. First factor or component accounts for 36.025% of the variability in all ten variables. Second component accounts for 29.404% of the variability in all ten variables. Third component accounts for 12.263% of the variability in all ten variables. Fourth component accounts for 6.215% of the variability in all ten variables. While it shows that the eigenvalue became less than 1 starting with the fourth component, the scree plot in Figure 13 starts to flatten out after the fourth component. Hence, middle school performance-engagement factors for Study 1 (until Study 3) consisted of four components or factors in modeling college outcomes.

Table 17

Total Variance Explained of PCA Components

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.602	36.025	36.025	3.602	36.025	36.025
2	2.940	29.404	65.429	2.940	29.404	65.429
3	1.226	12.263	77.692	1.226	12.263	77.692
4	0.621	6.215	83.907	0.621	6.215	83.907

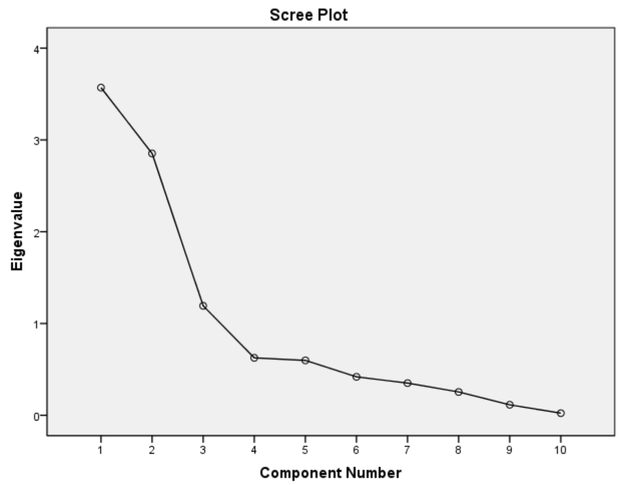


Figure 14. Scree plot of Principal Component Analysis results.

When loadings or correlations of less than 0.4 were excluded from Table 16, a four-component solution showed the following: knowledge, correctness and carelessness had high positive loadings on component 1, confusion had a moderate negative loading on component 1. This component can be labeled as *Aptitude*. Students who experienced greater amount of confusion would indicate having a lower *Aptitude*, while students having higher knowledge, carelessness, or correctness would indicate having higher *Aptitude*.

Component 2 had high negative loadings for number of actions and gaming the system, while moderate-to-strong positive loadings for correctness, boredom, confusion, and off-task behavior. This component can be labeled as *Disinterested Success* (or successful but disengaged). Students who had higher activity within ASSISTments and who gamed the system more had lower *Disinterested Success*, while students who had showed higher correctness, boredom, confusion, or off-task behavior had higher *Disinterested Success*. Such factor may be attributed to high ability students who used the system but showed lack of interest in undergoing the learning activities. This may be a sign of successful students who may find the material too easy and may get bored (San Pedro et al., 2013).

For component 3, frustration had a very positive loading, and boredom and confusion also had moderate positive loadings. This component can be labeled as *Negative Emotions*. Students who had higher occurrences of boredom, confusion or frustration, had higher *Negative Emotions*.

For component 4, engaged concentration had the sole loading above the cut-off. Hence this component can be labeled as *Engaged Concentration.* These defined components formed the structure of the middle school performance-engagement factors used in the creating the models for Study 1 – specifically, using middle school performance-engagement factors predicted by specific middle school variables (based on the structure in Figure 14), as the predictors of the college outcomes.

It is notable that across these four components or factors, a few middle school variables had cross loadings. Correctness had its highest loading from the *Aptitude* factor but had a cross-loading on the *Disinterested Success*. This makes sense as correctness attributes to the performance of students using ASSISTments. Boredom had a higher loading on *Disinterested Success*, but cross-loaded on *Negative Emotions* as well. Confusion had cross-loadings across *Aptitude*, *Disinterested Success* and *Negative Emotions*. The cross loadings for boredom and confusion across such factors demonstrate the varied relations of these academic emotions to learning activities – boredom can be evident to both successful students (finding the materials too easy) and unsuccessful students (after struggling with a difficult material), while confusion occurs in students who undergo an activity which may either be a sign of meaningful learning (resolving confusion and understanding the material) or poor learning (confusion is persistent or unable to resolve confusion).

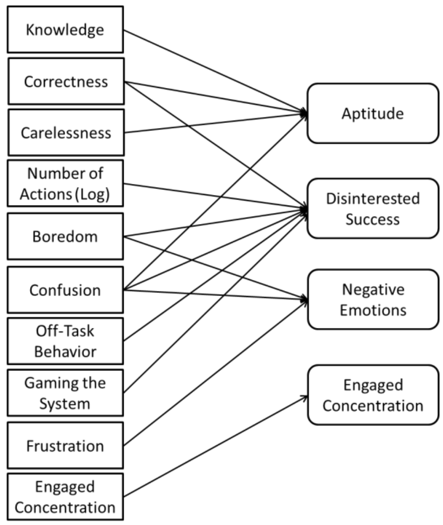


Figure 15. Structure of middle school performance-engagement factors based on PCA.

Structural Model of College Enrollment and Selectivity of College Attended

Figure 15 shows the multivariate model of college enrollment and selectivity of college attended created from the 7,636 students.

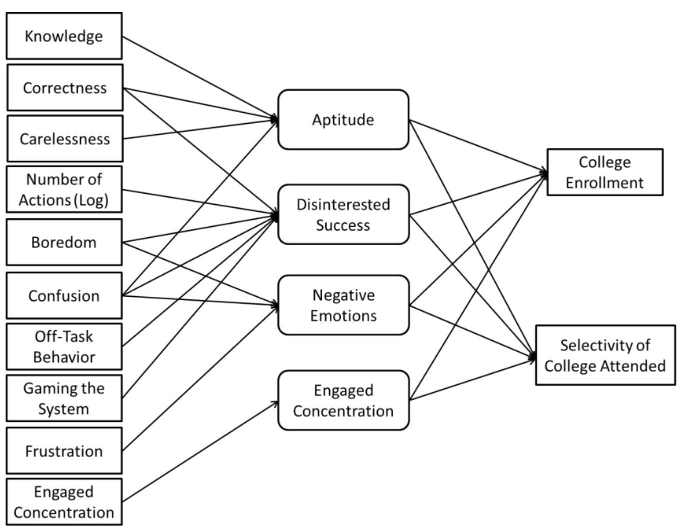


Figure 16. Model of college enrollment and selectivity of college attended using middle school performance-engagement factors.

As mentioned in Chapter Five, the college outcomes in Study 1 were modeled using both observed ("constrained") and unobserved ("unconstrained") representations of the middle school performance-engagement factors (middle layer in Figure 15), not only to compare the resulting estimates of these factors as predictors of the outcomes, but also to confirm the factor structure identified in the previous PCA. Table 18 summarizes the standardized model results of modeling college enrollment and selectivity of college attended using the four middle school performance-engagement factors as predictors. Table 18.a shows the model that used the component scores generated in PCA to represent middle school performance-engagement factors ("constrained"). Each of these middle school factors was then predicted by individual middle school variables as established by the factor structure in the previous PCA. Table 18.b shows the model that used middle school performance-engagement factors represented by unobservable, latent variables that were instead fit from individual middle school variables ("unconstrained") based again on the factor structure in the previous PCA.

For the model with constrained middle school performance-engagement factors in Table 18.a (LogLikelihood = -17435.615; AIC = 34951.231; BIC = 35228.856; Number of free parameters = 40), the significant relationships between individual middle school variables and the middle school factors confirm the loadings of the components from PCA (Table 16). For example, knowledge, carelessness and correctness had statistically significant positive coefficients, while confusion had a statistically significant negative coefficient for the *Aptitude* factor. The outcome of college enrollment was significantly predicted by *Aptitude*, *Disinterested Success*, *Negative Emotions* and *Engaged Concentration*. The likelihood of the student enrolling in college increases with a unit increase in *Aptitude* (β = 0.169, *SE* = 0.013, p<.001, Odds Ratio = 1.374), a unit increase in *Disinterested Success* (β = 0.085, *SE* = 0.012, p<.001, Odds Ratio =

1.173), a unit increase in *Engaged Concentration* (β = 0.089, *SE* = 0.013, *p*<.001, Odds Ratio = 1.183), or a unit decrease in *Negative Emotions* (β = -0.118, *SE* = 0.013, *p*<.001, Odds Ratio = 0.801), holding all other factors constant. Similarly, the outcome of selectivity of college attended was also significantly predicted by *Aptitude*, *Disinterested Success*, *Negative Emotions* and *Engaged Concentration*. The likelihood of the selectivity of college attended increasing a higher level increases with a unit increase in *Aptitude* (β = 0.309, *SE* = 0.012, *p*<.001, Odds Ratio = 1.849), a unit increase in *Disinterested Success* (β = 0.087, *SE* = 0.011, *p*<.001, Odds Ratio = 1.190), a unit increase in *Engaged Concentration* (β = 0.102, *SE* = 0.012, *p*<.001, Odds Ratio = 1.225), and a unit decrease in *Negative Emotions* (β = -0.154, *SE* = 0.012, *p*<.001, Odds Ratio = 0.735), holding all other factors constant.

For the model with "unconstrained" middle school performance-engagement factors in Table 18.b (LogLikelihood = -15600.042; AIC = 31254.085; BIC = 31441.482; Number of free parameters = 27), two middle school factors had deviations in their relationships with the individual middle school variables when compared to the factor loadings in the previous PCA. The factor *Aptitude* was positively associated with knowledge and correctness, but now negatively associated with carelessness and not significantly associated with confusion. The factor *Disinterested Success* now had a positive coefficient for number of actions and non-significant association with boredom. *Negative Emotions* was only significantly predicted by confusion and frustration, and *Engaged Concentration* was still defined by the individual middle school variable for engaged concentration. Despite the differences in the parameter estimates for some of the middle school factors, *Aptitude*, *Disinterested Success* and *Negative Emotions* still had significant associations with both college enrollment and selectivity of college attended. The unconstrained form of the factor *Engaged Concentration* was not significantly predictive anymore of college enrollment and selectivity of college attended.

The likelihood of the student enrolling in college increases with a unit increase in *Aptitude* (β = 0.174, *SE* = 0.019, *p*<.001, Odds Ratio = 1.040), a unit increase in *Disinterested Success* (β = 0.399, *SE* = 0.026, *p*<.001, Odds Ratio = 1.753), or a unit decrease in *Negative Emotions* (β = -0.108, *SE* = 0.021, *p*<.001, Odds Ratio = 0.841), holding all other factors constant. Similarly, the likelihood of the selectivity of college attended increasing a higher level increases with a unit increase in *Aptitude* (β = 0.079, *SE* = 0.019, *p*<.001, Odds Ratio = 1.019), a unit increase in *Disinterested Success* (β = 0.467, *SE* = 0.025, *p*<.001, Odds Ratio = 2.001), or a unit decrease in *Negative Emotions* (β = -0.157, *SE* = 0.022, *p*<.001, Odds Ratio = 0.767), holding all other factors constant.

Table 18

Standardized Model Results with (a) Constrained Middle School Performance-Engagement Factors; (b) Unconstrained Middle School Performance-Engagement Factors (College Outcomes: College Enrollment and Selectivity of College Attended)

	(a) Constrained Middle School Factors					*(b) Unconstrained Middle School Factors*				
	Estimate	*S.E.*	*p-value*	*Odds Ratio*	R^2	*Estimate*	*S.E.*	*p-value*	*Odds Ratio*	R^2
Aptitude ON					0.971					1.000[A]
Knowledge	0.295	0.012	<.001			1.098	0.420	0.009		
Carelessness	0.487	0.010	<.001			-2.025	0.343	<.001		
Correctness	0.228	0.005	<.001			0.115	0.010	<.001		
Confusion	-0.074	0.003	<.001			0.059	0.119	0.622		
Disinterested Success ON					0.964					1.000[A]
Correctness	0.166	0.004	<.001			0.717	0.055	<.001		
Number of Actions (log)	-0.339	0.004	<.001			0.805	0.051	<.001		
Boredom	0.097	0.004	<.001			-0.142	0.150	0.343		
Confusion	0.227	0.005	<.001			0.258	0.058	<.001		
Off-Task Behavior	0.243	0.004	<.001			0.129	0.037	0.001		
Gaming the System	-0.321	0.005	<.001			-0.375	0.048	<.001		
Negative Emotions ON					0.941					1.000[A]
Boredom	0.085	0.005	<.001			-0.682	0.485	0.160		
Confusion	0.234	0.005	<.001			0.816	0.067	<.001		

Frustration	0.828	0.003	<.001			0.816	0.067	<.001		
Engaged Concentration ON					0.738					1.000[A]
Engaged Concentration	0.859	0.003	<.001			1.000	<.001	na[B]		
College Enrollment ON					0.070					0.147
Aptitude	0.169	0.013	<.001	1.374		0.174	0.019	<.001	1.040	
Disinterested Success	0.085	0.012	<.001	1.173		0.399	0.026	<.001	1.753	
Negative Emotions	-0.118	0.013	<.001	0.801		-0.108	0.021	<.001	0.841	
Engaged Concentration	0.089	0.013	<.001	1.183		0.003	0.016	0.842	1.006	
Selectivity ON					0.171					0.233
Aptitude	0.309	0.012	<.001	1.849		0.079	0.019	<.001	1.019	
Disinterested Success	0.087	0.011	<.001	1.190		0.467	0.025	<.001	2.001	
Negative Emotions	-0.154	0.012	<.001	0.735		-0.157	0.022	<.001	0.767	
Engaged Concentration	0.102	0.012	<.001	1.225		-0.008	0.016	0.590	0.983	

Note: [A] Residual variance of 0 for latent factor; [B] undefined for fixed factor loading

Structural Model of Selectivity of College Attended Only

When students who did not enroll in college were dropped from the sample (leaving 4,131 students who did enroll in college), the sole college outcome of selectivity of college attended can be modeled (Figure 16) to predict the selectivity of the college the student enrolled in. Selectivity of college attended was again modeled with both constrained and unconstrained middle school performance-engagement factors.

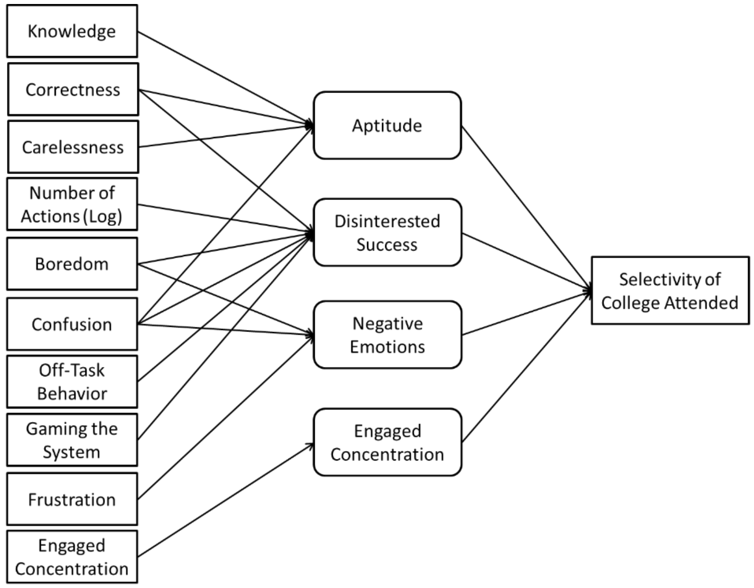


Figure 17. Model of selectivity of college attended using middle school performance-engagement factors.

The model results for selectivity of college attended in Table 19.a that used constrained middle school performance-engagement factors (LogLikelihood = -6082.685; AIC = 12235.370; BIC = 12456.790; Number of free parameters = 35) was similar to the model results in Table 18.a, where the significant relationships between individual middle school variables and the middle school factors confirm the loadings of the components from PCA. The middle school factors of *Aptitude*, *Disinterested Success* and *Engaged Concentration* had significant positive associations with selectivity of college attended, and *Negative Emotions* was negatively associated with selectivity of college attended. The likelihood of the selectivity of college attended increasing a higher level increases with a unit increase in *Aptitude* (β = 0.472, *SE* = 0.012, p<.001, Odds Ratio = 2.780), a unit increase in *Disinterested Success* (β = 0.102, *SE* = 0.013, p<.001, Odds Ratio = 1.253), a unit increase in *Engaged Concentration* (β = 0.106, *SE* = 0.013, p<.001, Odds Ratio = 1.256), or a unit decrease in *Negative Emotions* (β = -0.154, *SE* = 0.012, p<.001, Odds Ratio = 0.711), holding all other factors constant.

For the model with unconstrained middle school performance-engagement factors in Table 19.b (LogLikelihood = -6228.273; AIC = 12494.546; BIC = 12614.745; Number of free parameters = 19), two middle school factors resulted to losing significant associations with some of their individual middle school variables. *Aptitude* was only significantly predicted by knowledge, correctness and confusion. *Disinterested Success* was significantly predicted by correctness, number of actions, confusion and gaming. *Negative Emotions* and *Engaged Concentration* had the same significant associations with their respective middle school variables. These four middle school factors showed similar patterns as those in Table 19.a when predicting the college outcome of selectivity of college attended – the likelihood of the selectivity of college attended increasing a higher level increases with a unit increase in *Aptitude* (β = 0.372, *SE* = 0.023, *p*<.001, Odds Ratio = 1.140), a unit increase in *Disinterested Success* (β = 0.227, *SE* = 0.041, *p*<.001, Odds Ratio = 1.384), a unit increase in *Engaged Concentration* (β = 0.038, *SE* = 0.018, *p* = 0.034, Odds Ratio = 1.085), or a unit decrease in *Negative Emotions* (β = -0.168, *SE* = 0.048, *p* = 0.001, Odds Ratio = 0.861), holding all other factors constant.

Table 19

Standardized Model Results with (a) Constrained Middle School Performance-Engagement Factors; (b) Unconstrained Middle School Performance-Engagement Factors (College Outcome: Selectivity of College Attended Only)

	(a) Constrained Middle School Factors					*(b) Unconstrained Middle School Factors*				
	Estimate	*S.E.*	*p-value*	*Odds Ratio*	R^2	*Estimate*	*S.E.*	*p-value*	*Odds Ratio*	R^2
Aptitude ON					0.977					1.000[A]
Knowledge	0.294	0.015	<.001			0.163	0.066	0.013		
Carelessness	0.490	0.012	<.001			0.128	0.087	0.142		
Correctness	0.228	0.006	<.001			0.599	0.077	<.001		
Confusion	-0.079	0.003	<.001			-0.323	0.092	<.001		
Disinterested Success ON					0.970					1.000[A]
Correctness	0.167	0.005	<.001			0.664	0.081	<.001		
Number of Actions (log)	-0.356	0.005	<.001			0.664	0.081	<.001		
Boredom	0.081	0.006	<.001			0.311	0.177	0.079		
Confusion	0.261	0.007	<.001			0.664	0.081	<.001		

Off-Task Behavior	0.227	0.005	<.001			0.106	0.086	0.219		
Gaming the System	-0.298	0.006	<.001			-0.331	0.101	0.001		
Negative Emotions ON					0.942					1.000[A]
Boredom	0.067	0.008	<.001			0.413	0.003	<.001		
Confusion	0.243	0.008	<.001			0.413	0.004	<.001		
Frustration	0.831	0.005	<.001			0.413	0.005	<.001		
Engaged Concentration ON					0.731					1.000[A]
Engaged Concentration	0.855	0.004	<.001			1.000	<.001	na[B]		
Selectivity ON					0.309					0.294
Aptitude	0.472	0.012	<.001	2.780		0.372	0.023	<.001	1.140	
Disinterested Success	0.102	0.013	<.001	1.253		0.227	0.041	<.001	1.384	
Negative Emotions	-0.154	0.012	<.001	0.711		-0.168	0.048	0.001	0.861	
Engaged Concentration	0.106	0.013	<.001	1.256		0.038	0.018	0.034	1.085	

Note: [A] Residual variance of 0 for latent factor; [B] undefined

Study 2: Middle School → College Model (College STEM Major)

In Study 1, the middle school performance-engagement factors were defined and used to model the college outcomes of college enrollment and selectivity of college attended for 7,636 students. Study 2 used the same middle school performance-engagement factors for a smaller sample subset of 363 students to model another college outcome – namely, college major choice of a STEM major or non-STEM major (Figure 18.a) – using logistic regression. Study 2 also modeled this outcome from the individual middle school variables (Figure 18.b) with stepwise logistic regression to find combination of variables that have the best predictive power. Parameter estimates for both models have adjusted p-values to control for false discovery rate (FDR, Benjamini & Hochberg, 1995). Tables 20 and 21 show the correlations between middle school variables and performance-engagement factors, and outcome of college STEM major choice for the 363 students.

Table 20

Correlations between STEM Major Choice and Middle School Variables of Performance and Engagement for n = 363 students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Knowledge	1										
(2) Carelessness	.950**	1									
(3) Correctness	.840**	.684**	1								
(4) Number of Actions (log)	-.053	.094	-.307**	1							
(5) Boredom	-.234**	-.320**	.037	-.533**	1						
(6) Engaged Concentration	.068	.085	-.079	.486**	-.524**	1					
(7) Confusion	-.372**	-.431**	-.158**	-.452**	.753**	-.372**	1				
(8) Frustration	-.047	.002	-.172**	.048	.520**	-.109*	.322**	1			
(9) Off-Task Behavior	.230**	.163**	.299**	-.515**	.436**	-.409**	.195**	.151**	1		
(10) Gaming the System	-.427**	-.303**	-.694**	.569**	-.327**	.420**	-.273**	.134*	-.454**	1	
(11) STEM Major Choice	.269**	.239**	.308**	-.096	-.020	-.091	-.069	-.103	.041	-.279**	1

Table 21

Correlations between STEM Major Choice and Middle School Performance-Engagement Factors for n = 363 students

	(1)	(2)	(3)	(4)	(5)
(1) Aptitude	1				
(2) Disinterested Success	.194**	1			
(3) Negative Emotions	.039	.173**	1		
(4) Engaged Concentration	-.077	-.006	.048	1	
(5) STEM Major Choice	.277**	.166**	-.065	-.053	1

** - p < 0.05; ** - p< 0.001*

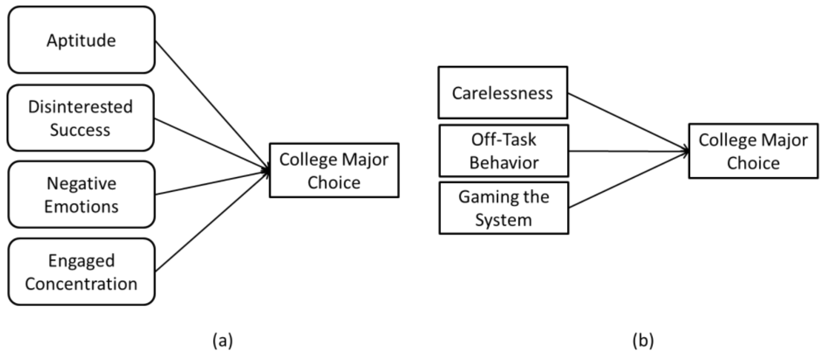


Figure 18. Model of college major choice: (a) Using middle school performance-engagement factors; (b) Using middle school variables.

Table 22 summarizes the model results for the 363 with college major information, using the middle school performance-engagement factors as predictors (Model 1) and the individual middle school variables as predictors (Model 2). For Model 1 (LogLikelihood = -231.729; AIC = 473.458; BIC = 492.930; Number of free parameters = 5), students were more likely to have enrolled in a STEM college major when they had higher *Aptitude* (β = 0.589, *SE* = 0.127, *p*<.001, *adjusted p* = 0.0125), higher *Disinterested Success* (β = 0.329, *SE* = 0.134, *p* = 0.014, *adjusted p* = 0.025), or lower *Negative Emotions* (β = -0.254, *SE* = 0.133, *p* = 0.056, *adjusted p* = 0.0375), holding all other factors constant. For Model 2 (LogLikelihood = -228.513; AIC = 465.027; BIC = 480.604; Number of free parameters = 4), the combination of middle school variables that best predicts college major choice (via stepwise logistic regression) consisted of gaming the system (β = -0.658, *SE* = 0.150, *p*<.001, *adjusted p* = 0.005), carelessness (β = 0.391, *SE* = 0.123, *p* = 0.001, *adjusted p* = 0.01), and off-task behavior (β = -0.273, *SE* = 0.134, *p* = 0.015, *adjusted p* = 0.015), where students who exhibited more off-task behavior or gaming the system were more likely to be in a non-STEM college major, while students who were more careless were more likely to be in a STEM college major.

Table 22

Standardized Model Results for Study 2 Predicting College Major Choice (n = 363 students)

(a) Middle School Factors as Predictors	*Estimate*	*S.E.*	*p-value*	*Odds Ratio*	R^2
College Major Choice ON					0.129
Aptitude	0.589	0.127	<.001†	1.802	
Disinterested Success	0.329	0.134	0.014†	1.389	
Negative Emotions	-0.254	0.133	0.056‡	0.775	
Engaged Concentration	-0.096	0.150	0.522	0.908	
Constant	-0.016	0.122	0.898	0.984	
(b)Middle School Variables as Predictors	*Estimate*	*S.E.*	*p-value*	*Odds Ratio*	R^2
College Major Choice ON					0.158
Gaming the System	-0.658	0.150	<.001†	0.518	
Carelessness	0.391	0.123	0.001†	1.479	
Off-Task Behavior	-0.273	0.134	0.041	0.761	
Constant	0.132	0.113	0.242	1.141	

† : significant after controlling FDR at 0.05 level (Benjamini & Hochberg, 1995);
‡: marginally significant after controlling for FDR at 0.05 level (Benjamini & Hochberg, 1995)

Study 3: Middle School → High School → College Model

Study 3 looked at possible high school variables that could mediate between the significant relationships found between middle school performance-engagement factors and college outcome for a sample of 282 students. The model parameter estimates presented here for Study 3 are non-standardized. Their corresponding standardized values are found in Appendix sections (Appendices E, F, G) . Zero-order correlations between the four middle school performance-engagement factors, four high school variables, and two college outcomes were first evaluated. Table 23 shows the strongest positive associations between the four middle school performance-engagement factors were found between Aptitude and Disinterested Success (r = 0.424, p<.001), and the strongest negative associations were between Aptitude and Negative Emotions (r = -0.303, p<.001). It is notable for the 282 students that Disinterested Success was only related to Aptitude and not Negative Emotions nor Engaged Concentration.

Table 23

Pearson Correlations between Middle School Performance-Engagement Factors for n = 282 students

	(1)	(2)	(3)	(4)
(1) Aptitude	1			
(2) Disinterested Success	.424**	1		
(3) Negative Emotions	-.303**	-.088	1	
(4) Engaged Concentration	.147*	-.111	-.205**	1

* - $p < 0.05$; ** - $p < 0.001$

Table 24 shows the correlations between the middle school performance-engagement factors and categorical high school variables and college outcomes. Between each of the middle school factors and each of the high school variables, only AP Math and AP Science were mostly correlated with the four middle school factors. AP Math was significantly correlated with only three middle school factors – with *Aptitude* (r = 0.572, p<.001), *Disinterested Success* (r = 0.253, p<.001), and *Negative Emotions* (r = 0.168, p = 0.005). AP Science was significantly correlated with all the four factors – with *Aptitude* (r = 0.480, p<.001),, *Disinterested Success* (r = 0.188, p = 0.002), *Negative Emotions* (r = -0.294, p<.001), and *Engaged Concentration* (r = 0.136, p = 0.023). Planned STEM major had only one significant correlation with a middle school factor – with *Aptitude* (r = 0.149, p = 0.013).

College enrollment was significantly correlated with three of the high school variables – AP Math (r = 0.185, p<.001), AP Science (r = 0.217, p<.001), and planned STEM major (r = 0.168, p<.001). Selectivity of college attended was significantly correlated with each of the high school variables – AP Math (r = 0.483, p<.001), AP Science (r = 0.504, p<.001), planned STEM major (r = 0.268, p<.001), and planned STEM career (r = 0.147, p = 0.013).

Table 24

Spearman Correlations between Middle School Performance-Engagement Factors, High School Variables and College Outcomes for n = 282 students

	Aptitude	*Disinterested Success*	*Negative Emotions*	*Engaged Concentration*	*AP Math*	*AP Science*	*Planned STEM Major*	*Planned STEM Career*
AP Math	.572**	.253**	-.168**	.086	1			
AP Science	.480**	.188**	-.294**	.136*	.541**	1		
Planned STEM Major	.149*	.046	-.012	.003	.285**	.299**	1	
Planned STEM Career	.109	.048	-.010	-.048	.143*	.136*	.541**	1
College Enrollment	.260**	.155**	-.051	.126*	.185**	.217**	.168**	.087
Selectivity of College Attended	.568**	.298**	-.221**	.145*	.483**	.504**	.268**	.147*

* - *p < 0.05;* ** - *p< 0.001*

Before conducting the processes for the mediational analyses between middle school factors, high school variables and college outcomes, significant relationships were first explored between a middle school factor and a college outcome for each group of a high school variable – for example, between students who took AP Math vs. students who did not take AP Math, between students who were interested in a STEM major vs students who were not interested in a STEM major, and so on (Table 25). This is visualized in Appendix D.

Table 25 shows the presence or absence of significant correlations between a college outcome and a middle school factor for each high school factor value. It can be seen that correlations between middle school *Aptitude* and college enrollment were significant among students who did not take AP Math or AP Science courses in high school, but not significant among students who took AP Math or AP Science. This may be counter-intuitive; perhaps this finding is specific to the 282 students of high ability who chose regular STEM courses in high school and performed well, and ended up attending college.

From Study 3 students, middle school *Disinterested Success* and *Engaged Concentration* were each found to be significantly correlated with college enrollment for those who were interested in high school to pursue a STEM major in college (Planned STEM major = 1), or a STEM career (Planned STEM career = 1), rather than for students who were interested in a non-STEM major or non-STEM career. This may be characteristic of high-achieving students when they were in middle school who were lackadaisical, but interested in the STEM domain when they were in high school.

Middle school *Negative Emotions* and *Engaged Concentration* were each significantly correlated with selectivity of college attended for students who took a regular math course in high school (AP Math = 0), but not significantly correlated for students who took AP math courses. This is indicative of these middle school factors potentially having a greater effect on selectivity for students who took less advanced math classes in high school (perhaps due to the pedagogical nature of these courses).

Middle school *Engaged Concentration* was significantly correlated with selectivity of college attended for students who were interested in high school to pursue a STEM major in college (Planned STEM major = 1), or a STEM career (Planned STEM career = 1), rather than for students who were interested in a non-STEM major or non-STEM career. This may be indicative of high-achieving students who have developed their interest in the STEM domain over the years.

Table 25

Spearman Correlations between Middle School Performance-Engagement Factors and College Outcomes between High School Variable Groups.

	AP Math = 0 (n=141)		AP Math = 1 (n=141)	
	College Enrollment	Selectivity of College Attended	College Enrollment	Selectivity of College Attended
Aptitude	.247**	.441**	.114	.385**
Disinterested Success	.141	.217**	.100	.236**
Negative Emotions	-.092	-.237**	.077	-.137
Engaged Concentration	.159	.182*	.012	.044
	AP Science = 0 (n=152)		AP Science = 1 (n=130)	
	College Enrollment	Selectivity of College Attended	College Enrollment	Selectivity of College Attended
Aptitude	.198*	.412**	.151	.429**
Disinterested Success	.121	.232**	.137	.290**
Negative Emotions	.013	-.143	-.015	-.029
Engaged Concentration	.138	.130	.003	.039
	Planned STEM Major = 0 (n=156)		Planned STEM Major = 1 (n=126)	
	College Enrollment	Selectivity of College Attended	College Enrollment	Selectivity of College Attended
Aptitude	.193*	.449**	.332**	.647**
Disinterested Success	.101	.224**	.251**	.386**
Negative Emotions	-.104	-.247**	.041	-.217*
Engaged Concentration	.094	.084	.180*	.222*
	Planned STEM Career = 0 (n=217)		Planned STEM Career (n=65)	
	College Enrollment	Selectivity of College Attended	College Enrollment	Selectivity of College Attended
Aptitude	.224**	.523**	.370**	.657**
Disinterested Success	.106	.253**	.337**	.431**
Negative Emotions	-.044	-.205**	.071	-.260*
Engaged Concentration	.077	.088	.354**	.360**

** - $p < 0.05$; ** - $p < 0.001$*

Creating a mediation model that used multiple middle school factors (*Aptitude, Disinterested Success, Negative Emotions, Engaged Concentration),* multiple high school mediators (AP Math, AP Science, planned STEM major, planned STEM career), and multiple college outcomes (college enrollment and selectivity of college attended) resulted in either non-convergence or non-significant relationships between middle school factors and college outcomes. Hence, Study 3 tested for mediational or indirect effects of individual high school variables in the relationships between individual middle school performance-engagement factors and individual college outcomes. To do this, zero-order relationships between the variables were first established (Figure 19) to see if: a middle school performance-engagement factor significantly predicts a college outcome (total effect) – coefficient ***c*** (Figure X.a); a middle school factor significantly predicts a high school mediator – coefficient ***a*** (Figure X.b); and lastly, a high school mediator, coefficient ***b,*** significantly predicts a college outcome controlling for middle school performance-engagement factor, coefficient ***c'*** (direct effect) (Figure X.c). If all these three relationships are found, the amount of mediation (indirect effect) is tested for significance, through Sobel test (MacKinnon, 2008). From this, the following outcomes occur: the presence of an indirect effect (significant mediation), the presence of full mediation where ***c*** is significant but ***c'*** is not, or the presence of partial mediation where ***c*** is significant and ***c'*** is non-zero and significant (Rucker, Preacher, Tormala, & Petty, 2011).

This process was conducted for each of the college outcome – college enrollment and selectivity of college attended.

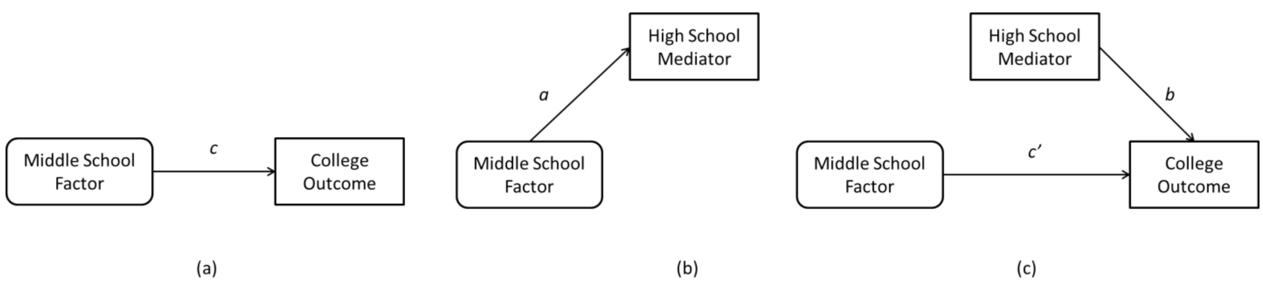


Figure 19. Steps for Mediation Analysis: (a) Middle school factor predicting college outcome; (b) Middle school factor predicting high school mediator; (c) High school mediator predicting college outcome controlling for middle school factor. Existing mediated or indirect effect is tested for significance after (a) to (b) using appropriate method.

Modeling College Enrollment with High School Mediating Variables

For the dichotomous college outcome of college enrollment, the first step was looking at its relationship with each of the middle school performance-engagement factors (total effect, coefficient ***c***), through binary logistic regression. Significant associations were found in Table 26 between college enrollment and *Aptitude* (***c*** = 0.684, *SE* = 0.164, *p*<.001, Odds Ratio = 1.982), college enrollment and *Disinterested Success* (***c*** = 0.485, *SE* = 0.219, *p* = 0.027, Odds Ratio = 1.624).

Table 26

Simple Logistic Regression Models of College Enrollment (Dichotomous Outcome) without High School Mediators for Study 3 (Significant relationship in bold)

	DV: College Enrollment			
Middle School Factors	*Estimate (**c**)*	*S.E.*	*p-value*	*Odds Ratio*
Aptitude	**0.684**	**0.164**	**<.001**	**1.982**
Constant	0.872	0.179	<.001	2.392
Disinterested Success	**0.485**	**0.219**	**0.027**	**1.624**
Constant	1.876	0.285	<.001	6.529
Negative Emotions	-0.270	0.160	0.092	0.764
Constant	1.193	0.178	<.001	3.296
Engaged Concentration	0.329	0.181	0.069	1.390
Constant	1.232	0.163	<.001	3.429

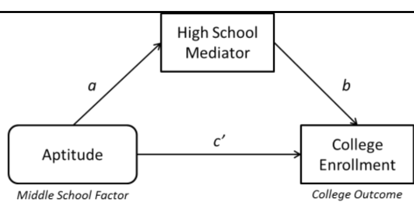
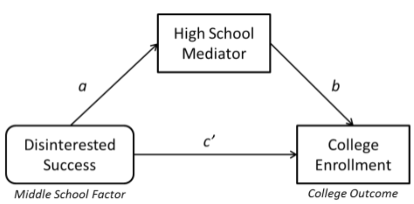
*Note: **c** coefficients are non-standardized binary logistic regression coefficients*

Succeeding steps in testing mediation were conducted between each of the middle school performance-engagement factors *Aptitude* and *Disinterested Success* and college enrollment – evaluating which among the high school variables of AP Math, AP Science, planned STEM major, and planned STEM career had a mediational or indirect effect between these relationships. *Negative Emotions* and *Engaged Concentration* did not have significant relationships with college enrollment, thus, I did not conduct mediational analyses that involved these middle school factors and college enrollment.

Table 27 shows non-standardized regression coefficients for four models of college enrollment predicted by *Aptitude* (mediated by each of the high school variables), and four models of college enrollment predicted by *Disinterested Success* (mediated by each of the high school variables). Appendix E shows the *Mplus* outputs of these models. With *Mplus* using MLR (maximum likelihood with robust standard errors) estimation in creating these models (as in Study 1), the high school variables were treated as continuous variables, creating OLS regression coefficients for coefficient ***a***, and binary logistic coefficients for coefficients ***b*** and ***c'*** (For comparison, Appendix E also includes an ML (maximum likelihood) estimation for these mediation models treating the high school variables as categorical mediators, and producing binary logistic coefficients for coefficients ***a***, ***b***, and ***c'***. It showed the same significant relationships as in Table 27, having the same ***b*** and ***c'*** coefficients, but different ***a*** coefficient.).

Table 27

Model Estimates of College Enrollment from Middle School Factors with Mediation of High School Variables (Significant mediations in bold)

Mediational Model	*High School Mediator*	*Middle School Factor → High School Mediator (**a**)*	*High School Mediator → College (**b**)*	*Middle School Factor → College Outcome (**c'**)*	*Indirect Effect (Sobel test statistic)*
	*AP Math*	0.292**	0.322	0.595*	0.852
	AP Science	**0.245****	**0.714^{+}**	**0.529^{+}**	**1.890^{+}**
	Planned STEM Major	0.077*	0.765*	0.649**	1.711
	Planned STEM Career	0.042	0.488	0.679**	0.948
	***AP Math***	**0.158****	**0.850***	**0.342**	**2.170***
	AP Science	**0.100***	**1.107***	**0.352**	**1.922^{+}**
	Planned STEM Major	0.024	0.880*	0.463*	0.586
	Planned STEM Career	0.021	0.545	0.474*	0.563

** p < 0.05; ** p< 0.001; + p is marginally significant*

*Note: **a** coefficients are OLS regression coefficients, **b** and **c'** coefficients are binary logistic regression coefficients*

For this case of a categorical outcome, a less straightforward, manual significance test of mediations or indirect effects was conducted by making the coefficients (***a***, ***b***, ***c'***, ***c***) comparable before conducting the Sobel test (MacKinnon & Dwyer, 1993; Muthen, 2011). Comparable coefficients are calculated by multiplying each coefficient by the standard deviation of the predictor variable in the equation and then dividing by the standard deviation of the outcome variable. Appendix H shows formulas used. I first discuss below the significant mediations found in Table 27.

Aptitude → High School Mediator → College Enrollment. Testing AP Science in high school as a mediator between *Aptitude* and college enrollment, *Aptitude* was found to significantly predict AP Science (***a*** = 0.245, *SE* = 0.022, *p*<.001). Using both *Aptitude* and AP Science as predictors, AP Science significantly predicted college enrollment (***b*** = 0.714, *SE* = 0.376, *p* = 0.057) after controlling for *Aptitude*, and direct path between *Aptitude* and college enrollment was still significant (***c'*** = 0.529, *SE* = 0.184, *p* = 0.004) after controlling for AP Science – consistent with partial mediation (Rucker, Preacher, Tormala, & Petty, 2011). Sobel test conducted showed the mediated or indirect effect by which *Aptitude* was associated with college enrollment through AP Science was marginally significant (*z* = 1.890, *p* = 0.06). Hence, the effects of *Aptitude* on taking AP Science in high school partially explained the effects of *Aptitude* on college enrollment.

Disinterested Success → High School Mediator → College Enrollment. Table 27 also shows that the significant relationship previously found between *Disinterested Success* and college enrollment for the 282 students can be explained by either the student's AP Math or AP Science. Looking at AP Math as a mediator between *Disinterested Success* and college enrollment, first, *Disinterested Success* significantly predicted AP Math (***a*** = 0.158, *SE* = 0.041, *p*<.001). Using both AP Math and *Disinterested* Success as predictors, AP Math significantly predicted college enrollment (***b*** = 0.850, *SE* = 0.325, *p* = 0.009) after controlling for *Disinterest Success*, but the direct path was not significant anymore between *Disinterested Success* and college enrollment (***c***' = 0.342, *SE* = 0.226, *p* = 0.129) after controlling for AP Math – consistent with full mediation (Rucker, Preacher, Tormala, & Petty, 2011). Sobel test conducted showed the mediated or indirect effect by which *Disinterested Success* was associated with college enrollment through AP Math was significant (*z* = 2.170, *p* = 0.032). Hence, the effects of middle school *Disinterested Success* on taking AP Math in high school may fully explain the effects of middle school *Disinterested Success* on college enrollment.

In testing a student's AP Science in high school as the mediator, *Disinterested Success* was first found to significantly predict AP Science (***a*** = 0.100, *SE* = 0.042, *p* = 0.018). AP Science significantly predicted college enrollment (***b*** = 1.107, *SE* = 0.342, *p* = 0.001), controlling for *Disinterested Success*. After controlling for AP Science, *Disinterested Success* was not a significant predictor of college enrollment anymore (***c'*** = 0.352, *SE* = 0.212, *p* = 0.097) – again consistent with full mediation (Rucker, Preacher, Tormala, & Petty, 2011). Sobel test conducted showed the mediated or indirect effect by which *Disinterested Success* was associated with college enrollment through AP Science was significant (*z* = 1.922, *p* = 0.05). Again, the effects of middle school *Disinterested Success* on taking AP Science in high school may fully explain the effects of middle school *Disinterested Success* on college enrollment.

Modeling Selectivity of College Attended with High School Mediating Variables

For the ordinal college outcome of selectivity of college attended, mediational analyses was first conducted by looking at its relationship with each of the middle school performance-engagement factors (total effect, coefficient ***c***) through ordered logistic regression. From Table 28, selectivity of college attended was significantly predicted by each of the middle school factors – *Aptitude* (***c*** = 1.279, *SE* = 0.134, *p*<.001), *Disinterested Success* (***c*** = 0.617, *SE* = 0.166,

p<.001), *Negative Emotions* ($\boldsymbol{c}$ = -0.556, *SE* = 0.118, p<.001) and *Engaged Concentration* ($\boldsymbol{c}$ = 0.291, *SE* = 0.107, p = 0.006).

Table 28

Simple Regression Models of Selectivity of College Attended (Ordinal Outcome) without High School Mediators for Study 3 (Significant relationship in bold)

	DV: Selectivity of College Attended (Logit)			
Middle School Factors	*Estimate (**c**)*	*S.E.*	*p-value*	*Odds Ratio*
Aptitude	**1.279**	**0.134**	**<.001**	**3.593**
Constant	**See Appendix F for thresholds between groups*			
Disinterested Success	**0.617**	**0.166**	**<.001**	**1.854**
Constant	**See Appendix F for thresholds between groups*			
Negative Emotions	**-0.556**	**0.118**	**<.001**	**0.574**
Constant	**See Appendix F for thresholds between groups*			
Engaged Concentration	**0.291**	**0.107**	**0.006**	**1.337**
Constant	**See Appendix F for thresholds between groups*			

*Note: **c** coefficients are non-standardized ordinal logistic regression coefficients*

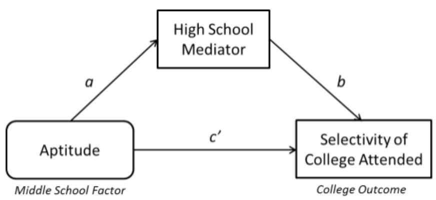
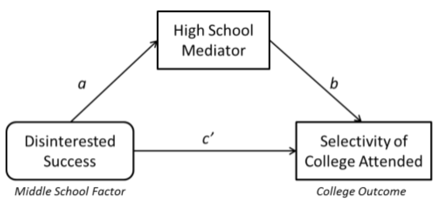
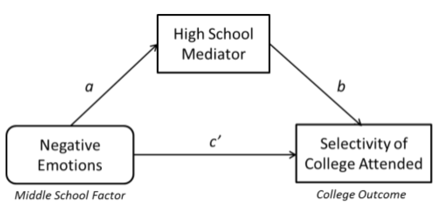
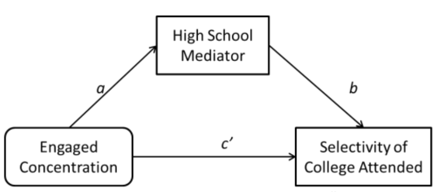
Succeeding steps in testing mediation were then conducted between selectivity of college attended and each of the middle school performance-engagement factors *Aptitude, Disinterested Success, Negative Emotions,* and *Engaged Concentration* – evaluating which among the high school variables of AP Math, AP Science, planned STEM major, and planned STEM career had a mediational or indirect effect between these relationships.

Table 29 shows non-standardized regression coefficients for four models of selectivity predicted by *Aptitude* (mediated by each of the high school variables), four models of selectivity predicted by *Disinterested Success* (mediated by each of the high school variables), four models of selectivity predicted by *Negative Emotions* (mediated by each of the high school variables), and four models of selectivity predicted by *Engaged Concentration* (mediated by each of the high school variables). Appendix G shows the *Mplus* outputs of these models. Again, with *Mplus* using MLR (maximum likelihood with robust standard errors) estimation in creating these models (as in Study 1), the high school variables were treated as continuous variables, creating OLS regression coefficients for coefficient ***a***, and ordinal logistic coefficients for coefficients ***b*** and ***c'*** (For comparison, Appendix G also includes an ML (maximum likelihood) estimation for these mediation models treating the high school variables as categorical mediators, and producing binary logistic coefficients for coefficients ***a***, and ordinal logistic coefficients for coefficients ***b*** and ***c'***. It showed the same significant relationships as in Table 29, having the same ***b*** and ***c'*** coefficients, but different ***a*** coefficient.).

Like in previous mediation models, a less straightforward, manual significance test of mediations or indirect effects was conducted by making the coefficients (***a***, ***b***, ***c'***, ***c***) comparable before conducting the Sobel test (Liu, Zhang, & Luo, 2015; MacKinnon & Dwyer, 1993). Comparable coefficients are calculated by multiplying each coefficient by the standard deviation of the predictor variable in the equation and then dividing by the standard deviation of the outcome variable. Appendix H shows formulas used. I first discuss below the significant mediations found in Table 29.

Table 29

Model Estimates of Selectivity of College Attended from Middle School Factors with Mediation of High School Variables (Significant mediations in bold)

Mediational Model	*High School Mediator*	*Middle School Factor → High School Mediator (**a**)*	*High School Mediator → College (**b**)*	*Middle School Factor → College Outcome (**c’**)*	*Indirect Effect (Sobel test statistic)*
	***AP Math***	**0.292****	**1.038****	**1.023****	**3.731****
	AP Science	**0.245****	**1.371****	**1.011****	**4.390****
	Planned STEM Major	**0.077***	**0.767****	**1.250****	**2.103***
	Planned STEM Career	0.042	0.412	1.265**	1.148
	***AP Math***	**0.158****	**1.840****	**0.404***	**3.408****
	AP Science	**0.100***	**2.047****	**0.490****	**2.273***
	Planned STEM Major	0.024	0.965**	0.600**	0.595
	Planned STEM Career	0.070	0.225*	0.355**	0.597
	***AP Math***	**-0.120****	**1.846****	**-0.448****	**-3.360****
	AP Science	**-0.197****	**1.966****	**-0.226**	**-4.943****
	Planned STEM Major	-0.020	1.006**	-0.575**	-0.601
	Planned STEM Career	0.004	0.662*	-0.566**	0.138
	*AP Math*	0.054	1.925**	0.242*	1.398
	AP Science	**0.096***	**2.071****	**0.171**	**2.675***
	Planned STEM Major	-0.008	1.002**	0.313*	-0.222
	Planned STEM Career	-0.022	0.653*	0.302*	0.726

** p < 0.05; ** p< 0.001; + p is marginally significant*
*Note: **a** coefficients are OLS regression coefficients, **b** and **c’** coefficients are ordered/ordinal logistic regression coefficients*

Aptitude → High School Mediator → Selectivity of College Attended. For the positive relationship between *Aptitude* and selectivity of the college attended, mediational or indirect effects of AP Math, AP Science and planned STEM major in high school were significant. Table 29 shows that *Aptitude* significantly predicted AP Math (***a*** = 0.292, *SE* = 0.020, *p*<.001). Using both *Aptitude* and AP Math as predictors of selectivity, AP Math significantly predicted selectivity after controlling for *Aptitude* (***b*** = 1.038, *SE* = 0.269, *p*<.001), while *Aptitude* and selectivity of the college attended were significantly related after controlling for AP Math (***c'*** = 1.023, *SE* = 0.148, *p*<.001). This significant relationship between *Aptitude* and selectivity after controlling AP Math is consistent with partial mediation (Rucker, Preacher, Tormala, & Petty, 2011). Sobel test conducted showed the mediated or indirect effect by which *Aptitude* was associated with selectivity through AP Math was significant (*z* = 3.731, *p*<.001).

Aptitude also significantly predicted AP Science (***a*** = 0.245, *SE* = 0.022, *p*<.001). Using both *Aptitude* and AP Science as predictors, AP Science significantly predicted selectivity of the college attended after controlling for *Aptitude* (***b*** = 1.371, *SE* = 0.287, *p*<.001) (Table 29), while *Aptitude* still significantly related to selectivity of the college attended after controlling for AP Science (***c'*** = 1.011, *SE* = 0.145, *p*<.001) – consistent with partial mediation. Sobel test conducted showed the mediated or indirect effect by which *Aptitude* was associated with selectivity through AP Science was significant (*z* = 4.390, *p*<.001).

Another mediator between *Aptitude* and selectivity of the college attended was planned STEM major (Table 29). *Aptitude* significantly predicted planned STEM major in high school (***a*** = 0.077, *SE* = 0.030, *p* = 0.010). Next, planned STEM major significantly predicted selectivity after controlling for *Aptitude* (***b*** = 0.767, *SE* = 0.209, *p*<.001), while *Aptitude* still significantly predicted selectivity of the college attended and controlling for planned STEM major (***c'*** = 1.250, *SE* = 0.139, *p*<.001). Sobel test conducted showed the mediated or indirect effect by which *Aptitude* was associated with selectivity through planned STEM major was significant (z = 2.103, p = 0.035).

With these three significant mediations, the positive effects of middle school *Aptitude* on selectivity of college attended may be partially explained by taking AP Math in high school, taking AP science in high school, or being interested in high school to pursue a STEM college major.

Disinterested Success → High School Mediator → Selectivity of College Attended. For the positive relationship between *Disinterested Success* and selectivity of the college attended, mediational or indirect effects of AP Math and AP Science in high school were significant. Table 29 shows that *Disinterested Success* significantly predicted AP Math (***a*** = 0.158, *SE* = 0.041, *p*<.001). With *Disinterested Success* and AP Math as predictors, selectivity was significant predicted by AP Math after controlling for AP Math (***b*** = 1.840, *SE* = 0.252, *p*<.001), while *Disinterested Success* still significantly predicted selectivity after controlling for AP Math (***c'*** = 0.404, *SE* = 0.154, *p* = 0.009). Sobel test conducted showed the mediated or indirect effect by which *Disinterested Success* was associated with selectivity through AP Math was significant (z = 3.408, *p*<.001).

Disinterested Success was also a significant predictor of AP Science (***a*** = 0.100, *SE* = 0.042, *p* = 0.018). With AP Science and *Disinterested Success* as predictors, selectivity was significantly predicted by AP Science after controlling for *Disinterested Success* (***b*** = 2.047, *SE* = 0.268, *p*<.001) (Table 29), while *Disinterested Success* was still significantly predicted selectivity after controlling for AP Science (***c'*** = 0.490, *SE* = 0.132, *p*<.001). Sobel test conducted showed the mediated or indirect effect by which *Disinterested Success* was associated with selectivity through AP Science was significant (z = 2.273, p = 0.023).

These mediation tests show that positive effects of middle school *Disinterested Success* on selectivity of college attended may be partially explained by taking AP Math or AP Science in high school.

Negative Emotions → High School Mediator → Selectivity of College Attended. Meanwhile, the negative significant relationship between *Negative Emotions* and selectivity of college attended was found to be partially mediated by either AP Math or fully mediated by AP Science in high school (Table 29). *Negative Emotions* significantly predicted AP Math (***a*** = -0.120, *SE* = 0.032, p<.001), while AP Math significantly predicted selectivity after controlling for *Negative Emotions* (***b*** = 1.846, *SE* = 0.244, p<.001), and *Negative Emotions* still significantly predicted selectivity after controlling for AP Math (***c'*** = -0.448, *SE* = 0.125, p<.001). Sobel test conducted showed the mediated or indirect effect by which *Negative Emotions* was associated with selectivity through AP Math was significant (z = -3.360, p<.001).

Negative Emotions also significantly predicted AP Science (***a*** = -0.197, *SE* = 0.028, p<.001), with AP Science significantly predicting selectivity after controlling for *Negative Emotions* (***b*** = 1.966, *SE* = 0.283, p<.001), and *Negative Emotions* not significantly predicting selectivity anymore after controlling for AP Science (***c'*** = -0.226, *SE* = 0.130, p = 0.083). Sobel test conducted showed the mediated or indirect effect by which *Negative Emotions* was associated with selectivity through AP Science was significant (z = -4.943, p<.001).

Hence, the negative effects of middle school *Negative Emotions* on selectivity of college attended may be explained by the type of math or science courses a student takes in high school.

Engaged Concentration → High School Mediator → Selectivity of College Attended. Lastly, a student's AP Science in high school fully mediated between the positive relationship between *Engaged Concentration* and selectivity of the college attended. *Engaged Concentration* significantly predicted AP Science (***a*** = 0.096, *SE* = 0.035, *p* = 0.006), AP Science significantly predicted selectivity after controlling for *Engaged Concentration* (***b*** = 2.071, *SE* = 0.265, *p*<.001) (Table 29), while *Engaged Concentration* was not significantly predictive anymore of selectivity after controlling for AP Science (***c'*** = 0.171, *SE* = 0.105, *p* = 0.102). Sobel test conducted showed the mediated or indirect effect by which *Engaged Concentration* was associated with selectivity through AP Science was significant (*z* = 2.675, *p* = 0.007). Hence, the effects of middle school *Engaged Concentration* on taking AP Science in high school can explain the effects of middle school *Aptitude* on selectivity of college attended.

CHAPTER VII.

DISCUSSION

This dissertation research conducted three studies to investigate how college attendance decisions can be attributed to factors within the student's educational experience as early as middle school. As Social Cognitive Career Theory proposes, both individual and environmental factors contribute to a student's learning experiences and greatly influence a student's academic and career choice (Lent, Brown, & Hackett, 1994; Lent, Brown, & Hackett, 2000). In this dissertation, the relationship between malleable and actionable factors known to occur during a student's learning experience and long-term college attendance outcomes were studied. This dissertation focused on actionable factors outside grades, tests and demographic information – namely, student knowledge, performance, academic emotions and behavior within a middle school computerized learning environment. This dissertation used secondary data from a student sample who used the ASSISTments system when they were in middle school, using existing fine-grained measures of knowledge, performance, academic emotions and behavior from their interaction data. This dissertation then examined the relationships between these middle school measures and college attendance outcomes of college enrollment, selectivity of college attended, and college major choice. Mediational analyses were conducted using data on students' high school AP Math, AP Science, planned STEM major, and planned STEM career.

Summary

Study 1 investigated the research question – ***Are student behavior, academic emotions and knowledge during middle school computer-based math learning predictive of college enrollment and selectivity of the college attended?*** This study identified middle school performance-engagement factors of *Aptitude, Disinterested Success*, *Negative Emotions* and

Engaged Concentration from the fine-grained measures of knowledge, correctness, number of actions, boredom, engaged concentration, confusion, frustration, off-task behavior, gaming the system and carelessness, and evaluated how these factors are predictive of the college outcomes of college enrollment and selectivity of college attended, using a sample of 7,636 students. The resulting models in Study 1 shed light on the potential impact of cognitive and non-cognitive factors experienced by middle school students in classrooms on their potential college attendance decisions. Academic emotions and engagement develop early in schooling, and become particularly prominent during the middle school years. Study 1 showed that together with a student's *Aptitude* and having high ability but showing disengaged behavior (*Disinterested Success*), *Negative Emotions* and *Engaged Concentration* experienced as early as middle school are predictive of eventual enrollment in college and the selectivity of college attended.

The positive relationship of middle school *Aptitude* with college enrollment and selectivity of college attended supports past research studies that used other indicators of academic performance (cf. Baron & Norman, 1992; Carnevale & Rose, 2003; Griffith & Rothstein, 2009), studies that identify college readiness to be linked to high performance during schooling (Roderick, Nagaoka, & Coca, 2009), as well as studies that predict that college enrollment is correlated with indicators of aptitude (Christensen, Melder, & Weisbrod, 1975; Eccles, Vida, & Barber, 2004). The positive relationship between middle school *Engaged Concentration* and college enrollment also aligns with previous works that show students who are more engaged in school tend to have higher academic motivation and achievement (Fredericks, Blumenfeld, & Paris, 2004; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013) that can lead to better preparation for college readiness (Balfanz, 2009; Conley, 2007).

While researchers have studied disengaged behavior of an intensity that leads to disciplinary referrals (Kellam, Ling, Meriska, Brown, & Ialongo, 1998; Reinke & Herman, 2002), the cognitive and non-cognitive factors studied in Study 1 may be more frequent, very mild in nature, and likely more actionable. This suggests that that in-the-moment interventions provided by software (or suggested by software to teachers) may have unexpectedly large effects, if they address negative affect and disengagement. Students who experience *Negative Emotions* can be properly supported in emotional self-regulation or by alternative instructional strategies or curriculum methods to address such emotions (e.g. boredom, confusion and frustration). Students who are *Disinterested Success* can be given content with greater novelty or challenge to increase their level of engagement and interest with the learning activities.

Study 2 investigated the research question – ***Are student behavior, academic emotions and knowledge during middle school computer-based math learning predictive of the pursuit or choice of a STEM college major once students are in college?*** This study analyzed the effects of the same fine-grained middle school measures of student knowledge, academic emotions and disengaged behavior, as well as effects of the middle school performance-engagement factors of *Aptitude, Disinterested Success, Negative Emotions,* and *Engaged Concentration* identified in Study 1 on the college outcome of college major choice (STEM or non-STEM major) available for 363 students.

Aptitude within middle school mathematics was positively associated with STEM major enrollment, a finding that aligns with studies that conceptualize high performance and developing aptitude during schooling as a sign of STEM major readiness and enrollment in STEM programs (Wang, 2012; Wang, 2013). *Disinterested Success* was also positively associated with pursuing a STEM college major, which aligns with this middle school factor's relationship with college enrollment and selectivity of college attended found in Study 1. *Negative Emotions* in middle school mathematics and its negative relationship with pursuing a STEM college major is a notable contrast to the previous work in (San Pedro, Ocumpaugh, Baker, & Heffernan, 2014) where affective states were not particularly strong predictors of STEM college major enrollment. This finding for *Negative Emotions* may be attributed to the similar associations that negative academic emotions such as boredom, confusion, and frustration, have with poorer learning outcomes (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013) and college enrollment (Study 1).

Study 2 also showed that the pursuit or choice of a STEM college major can also be influenced by fine-grained measures of middle school disengagement – carelessness, off-task behavior and gaming the system. This set of predictors is different from that in (San Pedro, Ocumpaugh, Baker, & Heffernan, 2014), where the feature selection was based on cross-validated A' metric. For Study 2, I used stepwise logistic regression. The positive association between carelessness in middle school and enrolling in a STEM college major may be non-intuitive, but this is in line with past results that careless errors are characteristic of more successful students (Clements, 1982). The disengaged behaviors of off-task behavior and gaming the system during middle school mathematics were found to be associated with not pursuing a STEM degree. These associations are not yet fully understood. Previous research has shown that that off-task behavior and gaming the system are associated with poorer learning (Cocea, Hershkovitz, & Baker, 2009), but they are also strong indicators of disengagement with mathematics, suggesting that students' lack of interest in STEM careers may manifest early. If these disengaged behaviors reduce the likelihood of pursuing a STEM major because they reduce learning, remediation or adjustments in instructional methods or curriculum may be relatively easy. If off-task behavior and gaming the system are instead an early indicator of lack of interest in STEM, remediation or alternative instructional strategies may be more difficult, but the information could still be used to provide actionable reports to teachers about their students' potential career interests.

Study 3 investigated the research question – ***How do high school course choices and interests in college majors and career during high school mediate between student behavior, academic emotions and knowledge in middle school computer-based math learning, and college attendance outcomes?*** This study evaluated potential mediational effects of high school factors of AP Math, AP Science, planned STEM major, and planned STEM career between the relationships that exist between middle school factors of *Aptitude, Disinterested Success, Negative Emotions,* and *Engaged Concentration* and college outcomes of college enrollment and selectivity of college attended (as established in Study 1), but for a smaller student sample of 282 students that also had high school data.

Study 3 showed that STEM course choices in high school (taking an AP or Honors math or science course) are strong mediators between middle school performance-engagement factors (*Aptitude, Disinterested Success, Negative Emotions* and *Engaged Concentration*) and college outcomes of college enrollment and selectivity of college attended. In addition, interest in pursuing a STEM college major during high school proved to mediate and partially explain the relationship between *Aptitude* and the college attendance outcomes.

These findings accords with SCCT-based theoretical accounts that experiences of mastery and motivation (as early as middle school) can drive future goals, interests and choices (in high school and beyond). A student who becomes disengaged during math learning is likely to dislike math (Baker, 2007; Baker, et al, 2008), and in turn have less interest in a math-related college major or career.

Cognitive and non-cognitive factors in middle school learning thus play a key role in the development of self-efficacy and vocational interests, and becoming prepared for college. And Study 3 shows how middle school learning and engagement can influence vocational interest and choices (pursuing college) by influencing student's self-efficacy for STEM courses (taking AP or Honors math or science courses in high school), as poor learning reduces self-efficacy and successful learning increases self-efficacy (cf. Bandura, 1997).

Findings in Study 3 support existing studies, with the middle school performance-engagement factors being both related to college attendance outcomes and high school course-taking. Eccles and Jacobs (1986) found that self-perceptions of math ability influenced math achievement and math course-taking plans, which aligns with middle school *Aptitude* and *Disinterested Success* being related to choice of AP Math or AP Science courses in high school.

Trusty (2002) found that college major choice (science or math majors) were very much related to course-taking, attitude, behavior, and self-perceptions of math ability in high school, as well academic performance in middle school. It was found that 8th-grade math test scores positively influenced math course-taking in high school for women, which in turn positively influenced later choice of science and math majors, while completing high school physics had a significant positive influence on choice of science and math majors for men (Trusty, 2002). Similarly, Trusty (2004) has also shown the effects of background variables (gender, race-ethnicity, SES) together with middle school reading and math abilities, and high school attendance behavior, positive school behavior and involvement in extracurricular activities on success in college.

It would be important to note that while middle school disengaged behavior such as off-task and gaming have been found to be related to negative attitude towards math, it would be helpful if support to address the occurrence of these behaviors would be further examined. Baker, Walonoski, and colleagues (2008) showed students who gamed the system with an education software for middle school mathematics (Cognitive Tutor) is both related to a student's negative attitude towards math as a subject, and to a student's negative attitude towards the software. Hence, the type of support a software provides to address students who game the system because they dislike math may be designed differently from the type of support to address students who game the system because they dislike the software (ex. lack of drive in learning math because of the difficulty of math content, compared to lack of drive in learning math because of poor design or presentation of the material in the math software).

As shown in this dissertation, examining how factors of student knowledge, academic emotions and behaviors of engagement or disengagement that are evident and influential in a student's experience with a learning environment, can potentially enrich the SCCT model by showing how these factors are also related to self-efficacy, interest and choice (Figure 20). Cognitive and non-cognitive factors within learning tasks (ex. with online learning environments) influence achievement and motivation to learn (Fredericks, Blumenfeld, & Paris, 2004; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013), and can manifest in the student's learning experiences defined by the learning strategies they use in classroom, or their behavioral and motivational engagement to pedagogy.

This dissertation is among the first studies that show the relationship of interaction-based measures of cognitive and non-cognitive constructs in learning and long-term outcomes.

Evaluating cognitive and non-cognitive aspects of learning in a dynamic way, through automated detectors or more immediate instructional feedback (either from educators or from the system itself), and relating this to long-term outcomes can be a starting point in re-evaluating and enriching factors to consider in counseling efforts.

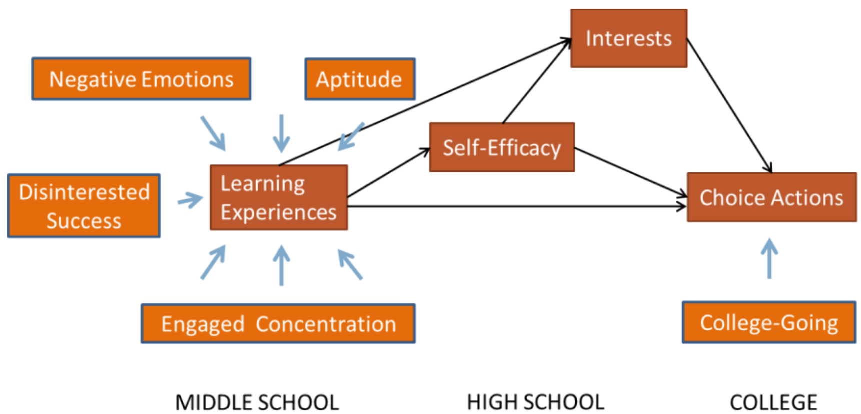


Figure 20. SCCT Model with Cognitive and Non-Cognitive Factors during Middle School Online Learning.

The middle school performance-engagement factors defined in this dissertation are among the many possible factors that can exist in the student-computer interaction with the ASSISTments system, and can also inform a student's learning experience in middle school. Studying cognitive and non-cognitive constructs within the context of learning with intelligent tutoring constructs offer both a lens about the student and the system itself. It can perhaps detect something about how students respond to online learning environment, leading to improvements in the pedagogical aspects of the system.

While the nature of the constructs explored in this dissertation suggests fail-soft interventions from educators and counselors may be feasible, it does not mean that they are less important compared to year-end grades or exam scores when evaluating student's academic success. This research shows how prevalent constructs of engagement and academic emotions are in learning activities (in and outside of digital learning environments), supporting current studies that show them to influence learning and achievement (McQuiggan, Mott, & Lester, 2008; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). As such, it would be important and beneficial for these factors need to be factored in instructional design.

It is also important to note that while the findings in this dissertation about middle school performance-engagement factors and their relations to college attendance may be indicative, they were evident and evaluated within the context of learning activities in middle school mathematics. And while it is possible that occurrences of these factors may overlap with other STEM domains (Alexander & Murphy, 1999; Dixon & Brown, 2012; Mayer & Wittrock, 1996; Rebello et al., 2007; Scherer & Beckmann, 2014), it is uncertain for non-STEM domains, as instructional content and learning strategies for non-STEM domain content would be very different from a STEM (ex. history, art, language). For example, we don’t know whether gaming behavior in a history class would lower odds of STEM major, lower odds of history major, or no effect at all.

Limitations

With the dissertation’s emphasis on fine-grained measures of cognitive and non-cognitive factors in a student’s middle school experience with an online learning environment, the lack of background variables from students (ex. gender, race, ethnicity, SES, etc) is a limitation in developing richer college attendance models geared towards counseling efforts. Hence, in terms of distinguishing between groups, the college attendance models presented in this research are limited and are not recommended to be used in this manner. However, in terms of evaluating at a more immediate frequency the student's learning experience – assessing the cognitive and non-cognitive constructs as early as middle school, it may be beneficial in guiding them to maintain a more productive and successful academic pipeline towards high school and college choices and outcomes.

It is also important to note that while Study 1, Study 2 and Study 3 conducted separate analyses and created separate models, there is potential for the model estimates in Study 2 and Study 3 to be susceptible to bias due to nonresponse errors – with Study 2 consisting of students who responded to post-high school survey, and Study 3 consisting of students who responded to the high school survey. It would be valuable for future work to be able to gather more individual differences for the student samples used in this dissertation to identify and adjust for any bias present, or increase the student sample size.

As mentioned in Chapter Six, Study 3 attempted to create a multiple-featured SEM that include multiple or all middle school performance-engagement factors, multiple or all high school variables, and multiple or all college outcomes, but this led to either non-significant relationships or non-convergence. With the substantial missing data problem from the Study 3 sample (when compared to Study 1 sample), data imputation would have led to less reliable estimates given the small number of students with observed high school variables and observed college major outcome.

It is important to note that while findings in this dissertation aim to inform designs in intervention and counseling efforts, decisions should be supplemented with sound judgment and expertise from educators in actual implementation of the intervention. The college attendance models presented in this research are not perfect and would benefit from being improved. Future work that can build on the college attendance models in this dissertation and can expound on the findings here may include gathering more background variables from the students, or other relevant measures of achievement, self-efficacy, and interests to improve the effect sizes of the college attendance models; evaluating the interactions between variables; or creating the models with a much bigger student sample size. Additional statistical analyses can also be conducted to further assess the stability and accuracy of the models. Power analysis can be conducted for Study 3 to address the potential bias in model estimates and of Type II error. Future work can also test these college attendance models for generalizability either with bootstrapping or cross-validation. Given further data variables, generalizability tests in the future could also be tested across gender, urbanicity, ethnicity, or across domains different from math – with students who used the ASSISTments software in a different subject domain such as Science and Technology or Engineering, or English Language Arts.

While findings in this dissertation show middle school cognitive and non-cognitive factors to be relevant early indicators of potential college attendance, successful entry to college should not be the only focus when providing classroom instruction and guidance as early as middle school. According to Campbell (1976), "*The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor*." Hence, quantitative assessments of the middle school cognitive and non-cognitive factors presented in this study should be supplemented with professional judgment from the educators when used in developing proper guidance and pathways for students towards academic success as they progress from one grade level to another. It must be supplemented with other measures of the student's learning experience, where teachers and counselors consult with one another to develop sound decisions in their counseling efforts with students – efforts that include how students can cope with their current learning experiences, not whether they will go to college or not. If a student's potential to go to college is solely judged by the middle school learning experience based only on any theoretical models, it undermines what can be changed in terms of pedagogy, learning strategies, or even parental involvement during the years between middle school and college. At best, the findings in this dissertation are a smaller but significant part of the span of educational experiences from K-12 to college.

Implications

Prior to graduating high school, students are faced with college-attendance choices. They explore options for a potential career, and consider which college or postsecondary institution they should attend. During K-12 learning, it is the educator's responsibility to guide students in discovering these options and help find a good fit for them. In career guidance counseling studies, questionnaire-based measures, which include self-efficacy and interest measures, are currently used to assess a student's likely career choice (cf. Betz, Borgen, & Harmon, 1996; Campbell, Hyne, & Nilsen, 1992) and attitudes toward career domains (Tapia & Marsh, 2004).

As established in this dissertation research, online learning environments create a valuable opportunity to keep students on a path toward college and prevent them from dropping out of the academic pipeline. One way to do this is by identifying richer measures within the students' learning experiences that current self-report measures may not capture. In assessing students' learning experiences as early as middle school—through academic emotions, and engaged and disengaged behavior—there is a potential for more effective interventions based on better information.

This dissertation investigated the learning mechanisms that students experience during their middle school and high school years and evaluated how they can be significant antecedents to their decisions to pursue a postsecondary education, forming a model of the trajectory from students' educational experiences starting in middle school, and the student's eventual path towards college. This research used a variety of data sources, most prominently from an online learning environment during middle school (used in their curriculum) that provide computer-assisted information. Data acquired from online learning environments allow researchers to computationally model and assess cognitive constructs such as learning, academic emotions and behavior, using current learning analytics or educational data mining methodologies. The goals and methods in this research took advantage of using educational data from online learning environments in analyzing long-term educational outcomes, one of the first studies that take advantage of that possibility in using these educational data sources.

This dissertation research can be expected to provide educators and career counselors with a new lens on how to develop counseling interventions, helping students interested in specific subject matter or postsecondary plans to sustain their interest in pursuing those goals (e.g. going to college, taking up biology, etc.). This study was conducted in the context of mathematics learning and mathematics careers, and the findings here have the potential to replicate in other domains. As online learning becomes more prevalent in K-12 education across the full range of subject domains, with more students using them, we will be able to identify and select students with special gifts in many areas, helping to track every student to career choices where they can be successful and contribute to society. Future work, beyond the scope of this dissertation research, includes the development of student assessment reports that can be used by teachers and guidance counselors. These reports can be based on this research's overall mediation model about a student's engagement, academic emotions and learning patterns, and progress in their potential to attend college.

REFERENCES

ACT, Inc. (2006). *Ready for college and ready for work: Same or different?* Retrieved from http://www.act.org/path/policy/pdf/ReadinessBrief.pdf

Alexander, P. A., & Murphy, P. K. (1999). Nurturing the seeds of transfer: A domain-specific perspective. *International journal of educational research*, 31(7), 561-576.

American School Counselor Association (ASCA) (2005). *The ASCA National Model: A framework for school counseling programs*. Alexandria, VA: Author.

Anderman, E. M., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of educational Research*, 64(2), 287-309.

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., Woolf, B.P. (2007) Repairing Disengagement with Non-Invasive Interventions. Proceedings of the 13th International Conference on Artificial Intelligence in Education, 195-202.

Avery, C., & Kane, T.J. (2004). Student perceptions of college opportunities. The Boston COACH program. In C. Hoxby (ed.). *College choices: The economics of where to go, when to go, and how to pay for it*. Chicago: University of Chicago Press

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

Baker, R. S., Corbett, A. T., & Aleven, V. (2008, January). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *Intelligent Tutoring Systems* (pp. 406-415). Springer Berlin Heidelberg.

Baker, R. S., Corbett, A. T., Gowda, S. M., Wagner, A. Z., MacLaren, B. A., Kauffman, L. R., ... & Giguere, S. (2010). Contextual slip and prediction of student performance after use of an intelligent tutor. In *User Modeling, Adaptation, and Personalization* (pp. 52-63). Springer Berlin Heidelberg.

Baker, R.S., Corbett, A.T., Koedinger, K.R., Evenson, S.E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., and Beck, J. (2006). Adapting to When Students Game an Intelligent Tutoring System. In Proceedings *ITS 2006*, 392- 401.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004, April). Off-task behavior in the cognitive tutor classroom: when students game the system. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 383-390). ACM.

Baker, R. S., 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., Gowda, S. M., Corbett, A. T., & Ocumpaugh, J. (2012). Towards automatically detecting whether student learning is shallow. In *Intelligent Tutoring Systems* (pp. 444-453). Springer Berlin Heidelberg.

Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., ... & Rossi, L. (2012). Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. *International Educational Data Mining Society*.

Baker, R., Siemens, G. (2014) Educational data mining and learning analytics. In Sawyer, K. (Ed.) *Cambridge Handbook of the Learning Sciences*: 2nd Edition, pp. 253-274.

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. *JEDM-Journal of Educational Data Mining*, 1(1), 3-17.

Balfanz, R. (2009). Putting middle grades students on the graduation path: A policy and practice brief. Baltimore, MD: *Everyone Graduates Center & Talent Development Middle Grades Program*. Johns Hopkins University.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, 42(4), 223-235.

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

Bandura, A. (1986). *Social foundations of thought and action* (pp. 5-107). Prentice Hall.: Englewood Cliffs, NJ.

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

Baron, J., & Norman, M. F. (1992). SATs Achievement Tests, and High-School Class Rank as Predictors of College Performance. Educational and Psychological Measurement, 52, 1047-1047.

Bartel, C.A., & Saavedra, R. (2000). The Collective Construction of Work Group Moods. *Administrative Science Quarterly*, 45 (2), 197-231.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Betz, N. E., Borgen, F. H., & Harmon, L. W. (1996). *Skills Confidence Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Betz, N. E., & Hackett, G. (1997). Applications of self-efficacy theory to the career assessment of women. *Journal of Career Assessment*, 5, 383-402.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child development*, 78(1), 246-263.

Blustein, David L. (1992) Applying current theory and research in career exploration to practice. *The Career Development Quarterly* 41.2: 174-183.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., & Zhao, W. (2015). Automatic detection of learning-centered affective states in the wild. *Proceedings of the 2015 International Conference on Intelligent User Interfaces*. New York, NY: ACM.

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton University Press.

Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research*, 103(3), 191-207.

Bowers, A. J., & Sprott, R. (2012). Why tenth graders fail to finish high school: A dropout typology latent class analysis. *Journal of Education for Students Placed at Risk* (JESPAR), 17(3), 129-148.

Byars-Winston, A. M., & Fouad, N. A. (2008). Math and science social cognitive variables in college students contributions of contextual factors in predicting goals. *Journal of Career Assessment*, 16(4), 425-440.

Cabrera, A. F., & La Nasa, S. M. (2000). Understanding the college-choice process. *New directions for institutional research*, *2000* (107), 5-22.

Cabrera, A. F., & La Nasa, S. M. (2001). On the path to college: Three critical tasks facing America's disadvantaged. *Research in Higher Education*, 42, 119-150.

Cabrera, A. F., La Nasa, S. M. and Burkum, K, R. (2001). *Pathways to a Four-Year Degree: The Higher Education Story of One Generation*. Center for the Study of Higher Education. Penn State University.

Camblin, S. (2003). *The middle grades: Putting all students on track for college*. Honolulu, HI: Pacific Resources for Education and Learning.

Campbell, D.P., Hyne, S.A., & Nilsen, D. (1992). *Manual for the Campbell Interest and Skill Survey* . Minneapolis, MN: National Computer Systems.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and program planning*, 2(1), 67-90.

Canfield, W. (2001). ALEKS: A Web-based intelligent tutoring system. *Mathematics and Computer Education*, 35(2), 152-158.

Carnevale, A. P., & Desrochers, D. M. (2003). *Standards for what?: The economic roots of K-16 reform.* Communication and Public Affairs, Office of Assessment, Equity, and Careers, Educational Testing Service.

Carnevale, A. P., & Rose, S. J. (2003). *Socioeconomic status, race/ethnicity, and selective college admissions*. New York: Century Foundation.

Carnevale, A. P., Smith, N., & Strohl, J. (2010). *Help wanted: Projections of job and education requirements through 2018*. Lumina Foundation.

Christensen, S., Melder, J., & Weisbrod, B. A. (1975). Factors affecting college attendance. *Journal of Human Resources*, 174-188.

Christenson, S. L., & Thurlow, M. L. (2004). School dropouts: Prevention considerations, interventions, and challenges. *Current Directions in Psychological Science*, 13(1), 36-39.

Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education*, 42(3), 309-328.

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

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.

College Division of Barron's Education Series (Ed.). (2012). *Barron's profiles of American colleges (30th ed.).* Hauppauge, NY: Barron's Educational Series, Inc.

Conley, D. T. (2007). *Redefining College Readiness*. Educational Policy Improvement Center.

Conley, D. T. (2008). *College knowledge: What it really takes for students to succeed and what we can do to get them ready*. John Wiley & Sons.

Conley, D. T. (2010). *College and career ready: Helping all students succeed beyond high school*. John Wiley & Sons.

Conley, D., Lombardi, A., Seburn, M., & McGaughy, C. (2009). *Formative Assessment for College Readiness: Measuring Skill and Growth in Five Key Cognitive Strategies Associated with Postsecondary Success.* Online Submission.

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: The Psychology of Optimal Experience*. Harper-Row, New York.

Cupani, M., & Pautassi, R. M. (2013). Predictive Contribution of Personality Traits in a Sociocognitive Model of Academic Performance in Mathematics. *Journal of Career Assessment*, 21(3), 395-413.

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. K., Lehman, B. Pekrun, R., & Graesser, A. C. (2014). Confusion Can be Beneficial For Learning. *Learning & Instruction*, 29(1), 153-170.

D'Mello, S. K., Taylor, R., & Graesser, A. C. (2007). Monitoring Affective Trajectories during Complex Learning. In D. S. McNamara & J. G. Trafton (Eds.), Proceedings of the 29th Annual Cognitive Science Society (pp. 203-208).

Daniels, L. M., Stupnisky, R. H., Pekrun, R., Haynes, T. L., Perry, R. P., & Newall, N. E. (2009). A longitudinal analysis of achievement goals: From affective antecedents to emotional effects and achievement outcomes. *Journal of Educational Psychology*, 101(4), 948.

Desmarais, M. C., & d Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1-2), 9-38.

Dixon, R. A., & Brown, R. A. (2012). Transfer of learning: Connecting concepts during problem solving. *Transfer*, 24(1).

Dowson, M., & McInerney, D. M. (2003). What do students say about their motivational goals?: Towards a more complex and dynamic perspective on student motivation. *Contemporary educational psychology*, 28(1), 91-113.

Dragon, T., Arroyo, I., Woolf, B. P., Burleson, W., El Kaliouby, R., & Eydgahi, H. (2008, January). Viewing student affect and learning through classroom observation and physical sensors. In Intelligent tutoring systems (pp. 29-39). Springer Berlin Heidelberg.

Drummond, J., & Litman, D. (2010, January). In the zone: Towards detecting student zoning out using supervised machine learning. In *Intelligent Tutoring Systems* (pp. 306-308). Springer Berlin Heidelberg.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child development*,82(1), 405-432.

Eccles, J. S., Vida, M. N., & Barber, B. (2004). The relation of early adolescents' college plans and both academic ability and task-value beliefs to subsequent college enrollment. *The Journal of Early Adolescence*, 24(1), 63-77.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance--*A Critical Literature Review. Consortium on Chicago School Research*. 1313 East 60th Street, Chicago, IL 60637.

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic bulletin & review*, 20(3), 586-592.

Fogarty, J., Baker, R. S., & Hudson, S. E. (2005, May). Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction. In *Proceedings of Graphics Interface 2005* (pp. 129-136). Canadian Human-Computer Communications Society.

Fouad, N. A., & Smith, P. L. (1996). A test of a social cognitive model for middle school students: Math and science. *Journal of Counseling Psychology*,43(3), 338.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*,74(1), 59-109.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement* (pp. 763-782). Springer US.

Gibbons, M. M., & Borders, L. D. (2010). Prospective First-Generation College Students: A Social-Cognitive Perspective. *The Career Development Quarterly*,58(3), 194-208.

Gibbons, M. M., Borders, L. D., Wiles, M. E., Stephan, J., & Davis, P. E. (2006). Career and college planning needs of ninth graders--as reported by ninth graders. *Professional School Counseling,* 10, 168-178.

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

Gong, Y., Beck, J. E., & Heffernan, N. T. (2010). Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures. In *Intelligent Tutoring Systems* (pp. 35-44). Springer Berlin Heidelberg.

Gore, P. A. (2006). Academic self-efficacy as a predictor of college outcomes: Two incremental validity studies. *Journal of Career Assessment*, 14(1), 92-115.

Gottschalk, P. (1997). Inequality, income growth, and mobility: The basic facts. *The Journal of Economic Perspectives*, 11(2), 21-40.

Griffith, A. L., & Rothstein, D. S. (2009). Can't get there from here: The decision to apply to a selective college. Economics of Education Review, 28(5), 620-628.

Hanley, J., & McNeil, B. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 143, 29-36.

Haveman, R. H., & Smeeding, T. M. (2006). The role of higher education in social mobility. *The Future of Children*, 16(2), 125-150.

Herr, E. L., & Niles, S. (1997). Perspectives on Career Assessment of Work-Bound Youth. *Journal of Career Assessment*, 5(2), 137-150.

Hershkovitz, A., Baker, R.S.J.d., Gobert, J., Wixon, M., Sao Pedro, M. (2013) Discovery with Models: A Case Study on Carelessness in Computer-based Science Inquiry. *American Behavioral Scientist, 57* (10), 1479-1498.

Hershkovitz, A., Baker, R. S., Gowda, S. M., & Corbett, A. T. (2013). Predicting Future Learning Better Using Quantitative Analysis of Moment-by-Moment Learning. *EDM*, 13, 74-81.

Holland, J. L. 1997. Making vocational choices: A theory of vocational personalities and work environments . Psychological Assessment Resources.

Hossler, D., Braxton, J., & Coopersmith, G. (1989). Understanding student college choice. *Higher education: Handbook of theory and research*, 5, 231-288.

Hoxby, C. (2009). The Changing Selectivity of American Colleges. *Journal of Economic Perspectives* 2009, 23 (4): 95-118.

Hoxby, C. M., & Avery, C. (2012). *The missing" one-offs": The hidden supply of high-achieving, low income students* (No. w18586). National Bureau of Economic Research.

Inman, W.I., & Mayes, L. (1999). The importance of being first generation community college students. *Community College Review*, 26(4), 3-22.

Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *Journal of Higher Education*, 861-885.

Jeni, L., Cohn, J. F., & De La Torre, F. (2013, September). Facing Imbalanced Data--Recommendations for the Use of Performance Metrics. In Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on (pp. 245-251). IEEE.

Johnson, P., Nichols, C. N., Buboltz, W. C., & Riedesel, B. (2002). Assessing a holistic trait and factor approach to career development of college students. *Journal of College Counseling*, 5(1), 4-14.

Junker, B. & Sijtsma, K. (2001). Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric Item Response Theory. *Applied Psychological Measurement*, 25(3), 258–272.

Kai, S., Paquette, L., Baker, R.S., Bosch, N., D'Mello, S., Ocumpaugh, J., Shute, V., Ventura, M. (2015) A Comparison of Face-based and Interaction-based Affect Detectors in Physics

Playground. *Proceedings of the 8th International Conference on Educational Data Mining* , 77-84

Karweit, N., & Slavin, R. E. (1982). Time-on-Task - Issues of Timing, Sampling, and Definition. *Journal of Educational Psychology*, 74(6), 844-851.

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Pschopathology*, 10, 165–185.

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*. Cambridge, MA: Cambridge University Press.

Kort, B., Reilly, R., Picard, R. (2001). An Affective Model Of Interplay Between Emotions And Learning: Reengineering Educational Pedagogy—Building A Learning Companion. *Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges*, Madison, Wisconsin: IEEE Computer Society, 43-48.

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. In *Affective Computing and Intelligent Interaction* (pp. 175-184). Springer Berlin Heidelberg.

Lemieux, T. (2006). Post-secondary education and increasing wage inequality(No. w12077). *National Bureau of Economic Research*.

Lent, R. W., & Brown, S. D. (2006). Integrating person and situation perspectives on work satisfaction: A social-cognitive view. *Journal of vocational behavior*, 69(2), 236-247.

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. (2000). Contextual supports and barriers to career choice: A social cognitive analysis. *Journal of Counseling Psychology*, 47(1), 36–49.

Lent, R. W., Hackett, G., & Brown, S. D. (1999). A social cognitive view of school-to-work transition. *The career development quarterly*, 47(4), 297-311.

Lent, R. W., Sheu, H., Singley, D., Schmidt, J. A., Schmidt, L. C., & Gloster, C. S. (2008). Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students. *Journal of Vocational Behavior*, 73, 328-335.

Linnenbrink, E. A., & Pintrich, P. R. (2003). The Role of Self-Efficacy Beliefs in Student Engagement and Learning in the Classroom. *Reading &Writing Quarterly*, 19(2), 119-137.

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. S. (2013). Sequences of frustration and confusion, and learning. In *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 114-120).

Liu, H., Zhang, Y., & Luo, F. (2015). Mediation Analysis for Ordinal Outcome Variables. In *Quantitative Psychology Research* (pp. 429-450). Springer International Publishing.

Luse, A., Rursch, J. A., & Jacobson, D. (2014). Utilizing Structural Equation Modeling and Social Cognitive Career Theory to Identify Factors in Choice of IT as a Major. *ACM Transactions on Computing Education (TOCE)*, 14(3), 19.

MacKinnon, D.P. (2008). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum Associates.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation review*, 17(2), 144-158.

Mann, S., & Robinson, A. (2009). Boredom in the lecture theatre: an investigation into the contributors, moderators and outcomes of boredom amongst university students. *British Educational Research Journal*, 35(2), 243-258.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. *Handbook of educational psychology*, 47-62.

McQuiggan, S. W., Mott, B. W., & Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User modeling and user-adapted interaction*, 18(1-2), 81-123.

Mentis, H.M., 2007. Memory of frustrating experiences. In: Nahl, D., Bilal, D. (Eds.), *Information and Emotion. Information Today*, Medford, NJ

Miller, W.L., Baker, R., Labrum, M., Petsche, K., & Wagner, A. (2014). Boredom Across Activites, and Across the Year, within Reasoning Mind. *KDD Workshop on Data Mining for Educational Assessment and Feedback (ASSESS 2014)*.

Muthén, B. (2011). Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. *Manuscript submitted for publication.*

Narciss, S. (2004). The impact of informative tutoring feedback and self-efficacy on motivation and achievement in concept learning. *Experimental Psychology*, 51, 214–228.

National Middle School Association (NMSA). (2002*). Research summary #12: Academic achievement*. 2002. Retrieved January 14, 2003, from www.nmsa.org/research/ressum12.htm

National Science Foundation, National Center for Science and Engineering Statistics (NCSES). 2013. Science and Engineering Degrees: 1966–2010. Detailed Statistical Tables NSF 13-327. Arlington, VA. Available at http://www.nsf.gov/statistics/nsf13327/

Neild, R. C. (2009). Falling off track during the transition to high school: What we know and what can be done. *The Future of Children*, 19(1), 53-76.

Nugent, G., Barker, B., Welch, G., Grandgenett, N., Wu, C., & Nelson, C. (2015). A model of factors contributing to STEM learning and career orientation. *International Journal of Science Education*, 37(7), 1067-1088.

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.

Ocumpaugh, J., Baker, R.S.J.d., & Rodrigo, M.M.T. (2012). *Baker-Rodrigo Observation Method Protocol (BROMP) 1.0. Training Manual version 1.0.* Technical Report. New York, NY: EdLab.

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

Pascarella, E. T., Pierson, C. T., Wolniak, G. C., & Terenzini, P. T. (2004). First-generation college students: Additional evidence on college experiences and outcomes. *Journal of Higher Education*, 249-284.

Pathways to College Network. (2004). *A shared agenda: A leadership challenge to improve college access and success*. Boston: Author. Available online at http://www.pathwaystocollege.net/pdf/SharedAgenda_FullReport.pdf

Pavlik Jr, P. I., Cen, H., & Koedinger, K. R. (2009). Performance Factors Analysis--A New Alternative to Knowledge Tracing. *Online Submission.*

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.

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., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282). Springer US.

Perrone, K,, Sedlacek, W, E,, & Alexander, C, M, (2001), Gender and ethnic differences in career goal attainment. *The Career Development Quarterly,* 50, 168-178.

Pike, G. R. 2006. Students' personality types, intended majors, and college expectations: Further evidence concerning psychological and sociological interpretations of Holland's theory. *Research in Higher Education*, 47, 7, 801-822.

Planalp, S., DeFrancisco, V.L., Rutherford, D. (1996). Varieties of Cues to Emotion in Naturally Occurring Settings. *Cognition and Emotion*, 10 (2), 137-153.

Porter, S. R., and Umbach, P. D. 2006. College major choice: An analysis of person–environment fit. Research in Higher Education, 47, 4, 429-449.

Quinlan, J.R. (1992). *C4.5: Programs for Machine Learning*, Morgan Kaufman.

Ramos-Sánchez, L., & Nichols, L. (2007). Self-efficacy of first-generation and non-first-generation college students: The relationship with academic performance and college adjustment. *Journal of College Counseling*, 10(1), 6-18.

Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., ... & Rasmussen, K. P. (2005). The Assistment project: Blending assessment and assisting. In *Proceedings of the 12th Annual Conference on Artificial Intelligence in Education* (pp. 555-562).

Reardon, S., Baker, R., & Klasik, D. (2012). *Race, income, and enrollment patterns in highly selective colleges, 1982-2004*. Center for Education Policy Analysis, Stanford University. Retrieved from http://cepa. stanford. edu/content/race-income-and-enrollmentpatterns-highly-selective-colleges-1982-2004.

Rebello, N. S., Cui, L., Bennett, A. G., Zollman, D. A., & Ozimek, D. J. (2007). Transfer of learning in problem solving in the context of mathematics and physics. *Learning to solve complex scientific problems*, 223-246.

Reinke, W. M., & Herman, K. C. (2002). Creating school environments that deter antisocial behaviours in youth. *Psychology in the Schools*, 39, 549–559.

Roderick, M., Coca, V., & Nagaoka, J. (2011). Potholes on the road to college high school effects in shaping urban students' participation in college application, four-year college enrollment, and college match. *Sociology of Education*, 84(3), 178-211.

Roderick, M., Nagaoka, J., & Coca, V. (2009). College readiness for all: The challenge for urban high schools. *The Future of Children*, 19(1), 185-210.

Roderick, M., Nagaoka, J., Coca, V., & Moeller, E. (2008). *From High School to the Future: Potholes on the Road to College*. Research Report. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.

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

Rojewski, J. W. (2002). Preparing the workforce of tomorrow: A conceptual framework for career and technical education. *Journal of Vocational Education Research*, 27(1), 7-35.

Rowe, J. P., McQuiggan, S. W., Robison, J. L., & Lester, J. C. (2009, July). Off-Task Behavior in Narrative-Centered Learning Environments. In *AIED* (pp. 99-106).

Rozin, P., & Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, 3(1), 68.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 5(6), 359-371.

Rupp, A. A. (2007, April). Unique characteristics of cognitive diagnosis models. Paper presented at the *Annual Meeting of National Council on Measurement in Education*, Chicago, IL.

Sabourin, J., Mott, B., & Lester, J. (2011). Modeling Learner Affect with Theoretically Grounded Dynamic Bayesian Networks. In *Proc. ACII 2011*, 286-295.

Sabourin, J., Rowe, J., Mott, B., Lester, J. (2011). When off-task is on-task: The affective role of off-task behavior in narrative-centered learning environments. In Proceedings of *Artificial Intelligence in Education 2011*, 534-536.

Sagen, H. B., Dallam, J. W., & Laverty, J. R. (2000). Effects of career preparation experiences on the initial employment success of college graduates. *Research in Higher Education*, 41(6), 753-767.

San Pedro, M. O., Baker, R. S., Bowers, A. J., & Heffernan, N. T. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In *Proceedings of the 6th international conference on educational data mining* (pp. 177-184).

San Pedro, M. O., Baker, R. S., Heffernan, N. T., & Ocumpaugh, J. L. (2015). Exploring college major choice and middle school student behavior, affect and learning: what happens to students who game the system? In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 36-40). ACM.

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

San Pedro, M.O., 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*, pp. 276-279.

Savalei, V. (2010). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods*, 15, 352-367.

Scherer, R., & Beckmann, J. F. (2014). The acquisition of problem solving competence: evidence from 41 countries that math and science education matters. *Large-scale Assessments in Education*, 2(1), 1-22.

Schmidt, W., Burroughs, N., Cogan, L., & Houang, R. (2011). Are College Rankings an Indicator of Quality Education? In *Forum on Public Policy Online* (Vol. 2011, No. 3). Oxford Round Table. 406 West Florida Avenue, Urbana, IL 61801.

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review*, 1, 173-208.

Shute, V. J., Moore, G. R., & Wang, L. (2015). Measuring problem solving skills in Plants vs. Zombies 2. *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*. Madrid, Spain.

Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, 32, 247-261.

Smith, M.S., Morgan, D., & Schacht. (2003). *The National Aeronautics and Space Administration's FY2004 budget request: Description, analysis, and issues for congress*. Washington, DC: Government Printing Office.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology* 1982 (pp.290-312).

Stine, D. D., & Matthews, C. M. (2009). *The US science and technology workforce*. Federal Publications, 644.

Tapia, M., & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, 8(2), 16.

The College Board. (2014). *College Majors and Career Search*. Retrieved from https://bigfuture.collegeboard.org

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.

Tobin, T., Sugai, G., & Colvin, G. (1996). Patterns in middle school discipline records. Journal of Emotional and Behavioral Disorders, 4(2), 82-94.

Tracey, T. J., Robbins, S. B., & Hofsess, C. D. (2005). Stability and change in interests: A longitudinal study of adolescents from grades 8 through 12. *Journal of Vocational Behavior*, 66(1), 1-25.

Trusty, J. (2002). Effects of High School Course-Taking and Other Variables on Choice of Science and Mathematics College Majors. *Journal of Counseling & Development*, 80(4), 464-474.

Trusty, J. (2004). *Effects of students' middle-school and high-school experiences on completion of the bachelor's degree* (Doctoral dissertation, University Park, PA).

Turner, S., & Lapan, R. T. (2002). Career Self-Efficacy and Perceptions of Parent Support in Adolescent Career Development. *The Career Development Quarterly*, 51(1), 44-55.

U.S. Department of Education. (1999). *Getting ready for college early: A handbook for parents of students in the middle and junior high school years*. Washington, DC: U.S. Government Printing Office.

U.S. Treasury Department. (2012). *New Report from Treasury, Education Departments: The Economic Case for Higher Education*. Washington, DC.

Valadez, J. R. (1998). Applying to college: Race, class, and gender differences. *Professional School Counseling*, 14-20.

Walonoski, J. A., & Heffernan, N. T. (2006, January). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. *In Intelligent Tutoring Systems* (pp. 382-391). Springer Berlin Heidelberg.

Wang, X. 2012. Modeling Student Choice of STEM Fields of Study: Testing a Conceptual Framework of Motivation, High School Learning, and Postsecondary Context of Support. WISCAPE Working Paper. Wisconsin Center for the Advancement of Postsecondary Education.

Wang, X. (2013). Why Students Choose STEM Majors Motivation, High School Learning, and Postsecondary Context of Support. *American Educational Research Journal*, 0002831213488622.

Wasson, A. S. (1981). Susceptibility to boredom and deviant behavior at school. *Psychological Reports*, 48(3), 901-902.

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, 45(5), 432-445.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

APPENDIX A

Models of Academic Emotions and Behaviors

Boredom Model for Students from Urban Schools (JRip)

```
(SumtimeTaken >= 103.422) => Bored=BORED (326.0/104.0)
(SumfrWorkingInSchool <= 1) and (Averagecorrect <= 0) => Bored=BORED (281.0/113.0)
 => Bored=NOT (641.0/228.0)

Number of Rules : 3
```

Confusion Model for Students from Urban Schools (J48)

```
Mincorrect <= 0
|   MaxtimeGreater5Secprev2wrong <= 0
|   |   Maxcorrect <= 0
|   |   |   MaxtotalFrPastWrongCount <= 0
|   |   |   |   Maxhint <= 0
|   |   |   |   |   SumconsecutiveErrorsInRow <= 0: NOT (119.0/45.0)
|   |   |   |   |   SumconsecutiveErrorsInRow > 0: CONFUSED (36.0/6.0)
|   |   |   |   Maxhint > 0: NOT (20.0)
|   |   |   MaxtotalFrPastWrongCount > 0
|   |   |   |   AveragetotalFrPastWrongCount <= 0.5: CONFUSED (87.0/12.0)
|   |   |   |   AveragetotalFrPastWrongCount > 0.5
|   |   |   |   |   MaxtotalFrPastWrongCount <= 4
|   |   |   |   |   |   Maxhint <= 0
|   |   |   |   |   |   |   SumconsecutiveErrorsInRow <= 0
|   |   |   |   |   |   |   |   MaxtotalFrPastWrongCount <= 2: CONFUSED (113.0/23.0)
|   |   |   |   |   |   |   |   MaxtotalFrPastWrongCount > 2: NOT (11.0)
|   |   |   |   |   |   |   SumconsecutiveErrorsInRow > 0
|   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount <= 2.75: NOT (33.0)
|   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount > 2.75
|   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount <= 3.571429
|   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow <= 23: NOT (6.0)
|   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow > 23: CONFUSED (15.0)
|   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount > 3.571429: CONFUSED (15.0)
|   |   |   |   |   |   Maxhint > 0: NOT (27.0)
|   |   |   |   |   MaxtotalFrPastWrongCount > 4
|   |   |   |   |   |   Maxhint <= 0: NOT (7.0)
|   |   |   |   |   |   Maxhint > 0
|   |   |   |   |   |   |   MaxtotalFrPastWrongCount <= 6: CONFUSED (45.0)
|   |   |   |   |   |   |   MaxtotalFrPastWrongCount > 6: NOT (3.0)
|   |   Maxcorrect > 0
|   |   |   MaxtotalFrPastWrongCount <= 2
|   |   |   |   Maxhint <= 0
|   |   |   |   |   MaxtotalFrPastWrongCount <= 0: CONFUSED (16.0/1.0)
|   |   |   |   |   MaxtotalFrPastWrongCount > 0
|   |   |   |   |   |   AveragetotalFrPastWrongCount <= 1.5: NOT (16.0)
|   |   |   |   |   |   AveragetotalFrPastWrongCount > 1.5: CONFUSED (17.0/2.0)
|   |   |   |   Maxhint > 0: NOT (5.0)
|   |   |   MaxtotalFrPastWrongCount > 2: NOT (20.0)
|   MaxtimeGreater5Secprev2wrong > 0
|   |   SumconsecutiveErrorsInRow <= 5
|   |   |   Maxhint <= 0
|   |   |   |   SumconsecutiveErrorsInRow <= 3
|   |   |   |   |   MaxtotalFrPastWrongCount <= 1
|   |   |   |   |   |   SumconsecutiveErrorsInRow <= 1: CONFUSED (30.0)
|   |   |   |   |   |   SumconsecutiveErrorsInRow > 1
|   |   |   |   |   |   |   SumconsecutiveErrorsInRow <= 2
|   |   |   |   |   |   |   |   MaxtotalFrPastWrongCount <= 0: CONFUSED (33.0/3.0)
|   |   |   |   |   |   |   |   MaxtotalFrPastWrongCount > 0: NOT (6.0)
|   |   |   |   |   |   |   SumconsecutiveErrorsInRow > 2: CONFUSED (52.0/7.0)
|   |   |   |   |   MaxtotalFrPastWrongCount > 1
|   |   |   |   |   |   Maxcorrect <= 0
|   |   |   |   |   |   |   SumconsecutiveErrorsInRow <= 2: CONFUSED (16.0/1.0)
|   |   |   |   |   |   |   SumconsecutiveErrorsInRow > 2: NOT (4.0)
|   |   |   |   |   |   Maxcorrect > 0: NOT (7.0)
|   |   |   |   SumconsecutiveErrorsInRow > 3: NOT (11.0)
|   |   |   Maxhint > 0
|   |   |   |   SumconsecutiveErrorsInRow <= 3
|   |   |   |   |   MaxtotalFrPastWrongCount <= 0: NOT (6.0)
|   |   |   |   |   MaxtotalFrPastWrongCount > 0
|   |   |   |   |   |   MaxtotalFrPastWrongCount <= 4
|   |   |   |   |   |   |   Maxcorrect <= 0: CONFUSED (19.0/4.0)
|   |   |   |   |   |   |   Maxcorrect > 0: NOT (3.0)
|   |   |   |   |   |   MaxtotalFrPastWrongCount > 4: CONFUSED (15.0)
|   |   |   |   SumconsecutiveErrorsInRow > 3: CONFUSED (141.0/6.0)
|   |   SumconsecutiveErrorsInRow > 5
|   |   |   AveragetimeGreater10SecAndNextActionRight <= 0.25
|   |   |   |   MaxtotalFrPastWrongCount <= 2
|   |   |   |   |   SumconsecutiveErrorsInRow <= 14
|   |   |   |   |   |   Maxhint <= 0
|   |   |   |   |   |   |   MaxtotalFrPastWrongCount <= 1: NOT (6.0)
|   |   |   |   |   |   |   MaxtotalFrPastWrongCount > 1: CONFUSED (15.0)
|   |   |   |   |   |   Maxhint > 0: NOT (19.0)
```

```
|   |   |   |   |   SumconsecutiveErrorsInRow > 14
|   |   |   |   |   |   SumconsecutiveErrorsInRow <= 21: CONFUSED (48.0/3.0)
|   |   |   |   |   |   SumconsecutiveErrorsInRow > 21
|   |   |   |   |   |   |   MaxtotalFrPastWrongCount <= 0: CONFUSED (16.0/1.0)
|   |   |   |   |   |   |   MaxtotalFrPastWrongCount > 0: NOT (9.0)
|   |   |   |   MaxtotalFrPastWrongCount > 2: NOT (18.0)
|   |   |   AveragetimeGreater10SecAndNextActionRight > 0.25: CONFUSED (15.0)
Mincorrect > 0
|   MaxtimeGreater5Secprev2wrong <= 0: NOT (339.0)
|   MaxtimeGreater5Secprev2wrong > 0
|   |   AveragetimeGreater10SecAndNextActionRight <= 0.5: NOT (45.0)
|   |   AveragetimeGreater10SecAndNextActionRight > 0.5
|   |   |   MaxtotalFrPastWrongCount <= 1
|   |   |   |   MaxtotalFrPastWrongCount <= 0: CONFUSED (21.0/6.0)
|   |   |   |   MaxtotalFrPastWrongCount > 0: NOT (11.0)
|   |   |   MaxtotalFrPastWrongCount > 1
|   |   |   |   MaxtotalFrPastWrongCount <= 2: CONFUSED (18.0/3.0)
|   |   |   |   MaxtotalFrPastWrongCount > 2: NOT (2.0)

Number of Leaves  :     46

Size of the tree :      91
```

Frustration Model for Students from Urban Schools (REP-Tree)

```
Averagecorrect < 0.58
|   AveragefrIsHelpRequest < 0.23
|   |   MaxhintTotal < 4.5
|   |   |   AveragetotalFrPastWrongCount < 0.42
|   |   |   |   MinhintTotal < 2.5 : NOT (48/0) [12/0]
|   |   |   |   MinhintTotal >= 2.5
|   |   |   |   |   AverageconsecutiveErrorsInRow < 1.43
|   |   |   |   |   |   Averagehint < 0.82
|   |   |   |   |   |   |   Averagehint < 0.4
|   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 0.5 : FRUSTRATED (74/11) [49/7]
|   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 0.5 : NOT (2/0) [1/0]
|   |   |   |   |   |   |   Averagehint >= 0.4 : FRUSTRATED (12/0) [3/0]
|   |   |   |   |   |   Averagehint >= 0.82 : NOT (2/0) [0/0]
|   |   |   |   |   AverageconsecutiveErrorsInRow >= 1.43 : NOT (6/0) [3/0]
|   |   |   AveragetotalFrPastWrongCount >= 0.42
|   |   |   |   AveragehintCount < 2.23
|   |   |   |   |   Minscaffold < 0.5
|   |   |   |   |   |   SumconsecutiveErrorsInRow < 5.5
|   |   |   |   |   |   |   AveragetotalFrPastWrongCount < 5.5
|   |   |   |   |   |   |   |   AveragehintCount < 1.25
|   |   |   |   |   |   |   |   |   MinhintTotal < 0.5
|   |   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount < 2.67
|   |   |   |   |   |   |   |   |   |   |   MaxhintTotal < 1
|   |   |   |   |   |   |   |   |   |   |   |   AverageconsecutiveErrorsInRow < 1.25
|   |   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 1.5 : FRUSTRATED (81/12) [46/10]
|   |   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 1.5 : NOT (2/0) [1/0]
|   |   |   |   |   |   |   |   |   |   |   |   AverageconsecutiveErrorsInRow >= 1.25 : FRUSTRATED (11/0) [4/0]
|   |   |   |   |   |   |   |   |   |   |   MaxhintTotal >= 1 : NOT (3/0) [1/0]
|   |   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount >= 2.67 : FRUSTRATED (11/0) [6/2]
|   |   |   |   |   |   |   |   |   MinhintTotal >= 0.5
|   |   |   |   |   |   |   |   |   |   Averagehint < 0.58
|   |   |   |   |   |   |   |   |   |   |   AveragetimeGreater10SecAndNextActionRight < 0.25
|   |   |   |   |   |   |   |   |   |   |   |   Averagecorrect < 0.13
|   |   |   |   |   |   |   |   |   |   |   |   |   AverageconsecutiveErrorsInRow < 3.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount < 0.83 : FRUSTRATED (12/1) [5/1]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount >= 0.83
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount < 2.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount < 1.75
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Averagehint < 0.25
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 0.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   MaxhintTotal < 2.5 : NOT (2/0) [2/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   MaxhintTotal >= 2.5 : FRUSTRATED (20/4) [15/1]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 0.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   MaxhintTotal < 2.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 2.5 : NOT (3/0) [4/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 2.5 : FRUSTRATED
(12/0) [3/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   MaxhintTotal >= 2.5 : NOT (8/0) [11/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Averagehint >= 0.25 : FRUSTRATED (7/0) [8/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount >= 1.75 : NOT (8/0) [4/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount >= 2.5 : FRUSTRATED (39/9) [19/4]
|   |   |   |   |   |   |   |   |   |   |   |   |   AverageconsecutiveErrorsInRow >= 3.5 : NOT (2/0) [3/0]
|   |   |   |   |   |   |   |   |   |   |   |   Averagecorrect >= 0.13 : NOT (10/0) [2/0]
|   |   |   |   |   |   |   |   |   |   |   AveragetimeGreater10SecAndNextActionRight >= 0.25 : FRUSTRATED (23/3) [14/4]
|   |   |   |   |   |   |   |   |   |   Averagehint >= 0.58
|   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 1.5 : NOT (2/0) [1/0]
|   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 1.5
|   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 3.5 : FRUSTRATED (20/0) [10/0]
|   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 3.5
|   |   |   |   |   |   |   |   |   |   |   |   |   MaxhintTotal < 3.5 : NOT (2/0) [1/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   MaxhintTotal >= 3.5 : FRUSTRATED (12/1) [5/1]
|   |   |   |   |   |   |   |   AveragehintCount >= 1.25 : NOT (11/0) [2/0]
|   |   |   |   |   |   |   AveragetotalFrPastWrongCount >= 5.5
|   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 1 : NOT (2/0) [1/0]
|   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 1 : FRUSTRATED (33/0) [13/1]
|   |   |   |   |   |   SumconsecutiveErrorsInRow >= 5.5
|   |   |   |   |   |   |   AveragetotalFrPastWrongCount < 6.95
|   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount < 0.95 : NOT (3/0) [0/0]
```

```
|   |   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount >= 0.95
|   |   |   |   |   |   |   |   |   |   Averagehint < 0.78
|   |   |   |   |   |   |   |   |   |   |   AverageconsecutiveErrorsInRow < 6.75 : FRUSTRATED (103/1) [53/5]
|   |   |   |   |   |   |   |   |   |   |   AverageconsecutiveErrorsInRow >= 6.75
|   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 15 : NOT (3/0) [1/0]
|   |   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 15 : FRUSTRATED (55/3) [26/3]
|   |   |   |   |   |   |   |   |   |   Averagehint >= 0.78
|   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow < 6.5 : FRUSTRATED (10/0) [5/0]
|   |   |   |   |   |   |   |   |   |   |   SumconsecutiveErrorsInRow >= 6.5 : NOT (6/0) [3/0]
|   |   |   |   |   |   |   |   AveragetotalFrPastWrongCount >= 6.95 : NOT (4/0) [1/0]
|   |   |   |   |   |   Minscaffold >= 0.5 : NOT (6/0) [3/0]
|   |   |   |   AveragehintCount >= 2.23
|   |   |   |   |   AveragehintCount < 3.33 : NOT (16/0) [3/0]
|   |   |   |   |   AveragehintCount >= 3.33 : FRUSTRATED (10/1) [7/1]
|   |   MaxhintTotal >= 4.5 : NOT (13/0) [11/0]
|   AveragefrIsHelpRequest >= 0.23
|   |   MaxhintTotal < 4.5 : NOT (38/0) [11/0]
|   |   MaxhintTotal >= 4.5 : FRUSTRATED (9/1) [10/3]
Averagecorrect >= 0.58
|   MinhintTotal < 3.5
|   |   MaxhintTotal < 1.5
|   |   |   MaxhintTotal < 0.5
|   |   |   |   AveragetotalFrPastWrongCount < 5 : NOT (103/8) [61/7]
|   |   |   |   AveragetotalFrPastWrongCount >= 5 : FRUSTRATED (11/1) [5/0]
|   |   |   MaxhintTotal >= 0.5 : FRUSTRATED (10/0) [5/0]
|   |   MaxhintTotal >= 1.5 : NOT (126/0) [49/0]
|   MinhintTotal >= 3.5
|   |   MaxhintTotal < 4.5
|   |   |   AveragehintCount < 1.5
|   |   |   |   AveragehintCount < 0.5 : NOT (32/12) [21/3]
|   |   |   |   AveragehintCount >= 0.5 : FRUSTRATED (16/3) [3/1]
|   |   |   AveragehintCount >= 1.5 : NOT (9/0) [1/0]
|   |   MaxhintTotal >= 4.5 : NOT (17/0) [12/0]

Size of the tree : 101
```

Off-task Model for Students from Urban and Suburban Schools (REP-Tree)

```
SumfrWorkingInSchool < 2.5
|   MintimeTaken < 1.84 : OFF TASK (34/0) [14/3]
|   MintimeTaken >= 1.84
|   |   SumtimeTaken < 97.12
|   |   |   SumtotalFrAttempted < 33.5
|   |   |   |   Maxcorrect < 0.5
|   |   |   |   |   SumtotalFrAttempted < 9.5 : OFF TASK (21/2) [10/3]
|   |   |   |   |   SumtotalFrAttempted >= 9.5
|   |   |   |   |   |   MintimeTaken < 7.01 : OFF TASK (8/2) [4/0]
|   |   |   |   |   |   MintimeTaken >= 7.01
|   |   |   |   |   |   |   SumtimeTaken < 62.06
|   |   |   |   |   |   |   |   MintimeTaken < 14.22
|   |   |   |   |   |   |   |   |   SumtimeTaken < 53.73
|   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.64 : OFF TASK (4/1) [1/0]
|   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.64 : NOT (2/0) [1/0]
|   |   |   |   |   |   |   |   |   SumtimeTaken >= 53.73 : NOT (3/0) [0/0]
|   |   |   |   |   |   |   |   MintimeTaken >= 14.22 : NOT (5/0) [4/2]
|   |   |   |   |   |   |   SumtimeTaken >= 62.06 : OFF TASK (5/1) [3/0]
|   |   |   |   Maxcorrect >= 0.5
|   |   |   |   |   SumtotalFrPercentPastWrong < 0.35
|   |   |   |   |   |   SumtimeTaken < 94.95
|   |   |   |   |   |   |   AverageendsWithAutoScaffolding < 0.25
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.31
|   |   |   |   |   |   |   |   |   SumtimeTaken < 29.02 : OFF TASK (29/12) [6/2]
|   |   |   |   |   |   |   |   |   SumtimeTaken >= 29.02
|   |   |   |   |   |   |   |   |   |   SumtimeTaken < 34.97 : NOT (5/0) [1/1]
|   |   |   |   |   |   |   |   |   |   SumtimeTaken >= 34.97
|   |   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted < 31.5
|   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted < 1.5 : OFF TASK (4/1) [2/0]
|   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted >= 1.5 : NOT (38/12) [19/4]
|   |   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted >= 31.5 : OFF TASK (3/0) [0/0]
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.31 : OFF TASK (6/1) [1/1]
|   |   |   |   |   |   |   AverageendsWithAutoScaffolding >= 0.25 : NOT (4/0) [2/1]
|   |   |   |   |   |   SumtimeTaken >= 94.95 : NOT (4/0) [0/0]
|   |   |   |   |   SumtotalFrPercentPastWrong >= 0.35 : NOT (33/5) [16/5]
|   |   |   SumtotalFrAttempted >= 33.5 : NOT (83/17) [34/10]
|   |   SumtimeTaken >= 97.12
|   |   |   MintimeTaken < 129.75
|   |   |   |   AveragefrPast5WrongCount < 1.25
|   |   |   |   |   SumtotalFrPercentPastWrong < 0.41
|   |   |   |   |   |   MintimeTaken < 12.04 : OFF TASK (14/1) [6/1]
|   |   |   |   |   |   MintimeTaken >= 12.04
|   |   |   |   |   |   |   SumtimeTaken < 251.17
|   |   |   |   |   |   |   |   SumtimeTaken < 217.37
|   |   |   |   |   |   |   |   |   MintimeTaken < 51.39
|   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted < 52.5
|   |   |   |   |   |   |   |   |   |   |   MintimeTaken < 23.76 : NOT (7/3) [5/1]
|   |   |   |   |   |   |   |   |   |   |   MintimeTaken >= 23.76
|   |   |   |   |   |   |   |   |   |   |   |   MintimeTaken < 35.51 : OFF TASK (9/0) [9/4]
|   |   |   |   |   |   |   |   |   |   |   |   MintimeTaken >= 35.51
|   |   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted < 5 : OFF TASK (5/0) [2/1]
|   |   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted >= 5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted < 10 : NOT (2/0) [3/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted >= 10 : OFF TASK (6/1) [2/1]
|   |   |   |   |   |   |   |   |   |   SumtotalFrAttempted >= 52.5 : NOT (3/0) [5/2]
```

```
|   |   |   |   |   |   |   |   |   MintimeTaken >= 51.39
|   |   |   |   |   |   |   |   |   |   MintimeTaken < 53.97 : NOT (5/0) [0/0]
|   |   |   |   |   |   |   |   |   |   MintimeTaken >= 53.97
|   |   |   |   |   |   |   |   |   |   |   SumtimeTaken < 192.98 : OFF TASK (16/7) [3/1]
|   |   |   |   |   |   |   |   |   |   |   SumtimeTaken >= 192.98 : NOT (2/0) [4/2]
|   |   |   |   |   |   |   |   SumtimeTaken >= 217.37 : OFF TASK (5/0) [3/1]
|   |   |   |   |   |   |   SumtimeTaken >= 251.17
|   |   |   |   |   |   |   |   SumtimeTaken < 294.07 : NOT (5/0) [0/0]
|   |   |   |   |   |   |   |   SumtimeTaken >= 294.07
|   |   |   |   |   |   |   |   |   SumtotalFrAttempted < 23 : OFF TASK (5/4) [4/0]
|   |   |   |   |   |   |   |   |   SumtotalFrAttempted >= 23 : OFF TASK (4/1) [3/0]
|   |   |   |   |   SumtotalFrPercentPastWrong >= 0.41 : NOT (13/3) [9/0]
|   |   |   |   AveragefrPast5WrongCount >= 1.25 : OFF TASK (14/2) [8/4]
|   |   |   MintimeTaken >= 129.75 : OFF TASK (26/3) [22/1]
SumfrWorkingInSchool >= 2.5 : NOT (125/11) [73/7]

Size of the tree : 71
```

Engaged Concentration Model for Students from Suburban Schools (J48)

```
MinfrWorkingInSchool <= 0
|   MinhintTotal <= 0: CONCENTRATING (1594.0/595.0)
|   MinhintTotal > 0
|   |   SumhintTotal <= 16
|   |   |   SumhintTotal <= 3
|   |   |   |   MinhintTotal <= 2
|   |   |   |   |   SumhintTotal <= 2: NOT (36.0/8.0)
|   |   |   |   |   SumhintTotal > 2: CONCENTRATING (4.0)
|   |   |   |   MinhintTotal > 2: NOT (92.0/15.0)
|   |   |   SumhintTotal > 3
|   |   |   |   MinhintTotal <= 1: NOT (23.0/2.0)
|   |   |   |   MinhintTotal > 1
|   |   |   |   |   SumhintTotal <= 14
|   |   |   |   |   |   SumfrWorkingInSchool <= 1
|   |   |   |   |   |   |   SumhintTotal <= 10
|   |   |   |   |   |   |   |   SumhintTotal <= 6
|   |   |   |   |   |   |   |   |   MinhintTotal <= 3: CONCENTRATING (11.0)
|   |   |   |   |   |   |   |   |   MinhintTotal > 3: NOT (23.0/9.0)
|   |   |   |   |   |   |   |   SumhintTotal > 6: NOT (20.0/6.0)
|   |   |   |   |   |   |   SumhintTotal > 10: CONCENTRATING (7.0)
|   |   |   |   |   |   SumfrWorkingInSchool > 1: NOT (8.0/1.0)
|   |   |   |   |   SumhintTotal > 14: NOT (15.0/1.0)
|   |   SumhintTotal > 16
|   |   |   SumhintTotal <= 18: NOT (9.0/2.0)
|   |   |   SumhintTotal > 18: CONCENTRATING (11.0)
MinfrWorkingInSchool > 0
|   SumfrWorkingInSchool <= 1
|   |   MinsumTime3SDWhen3RowRight <= 2.049127
|   |   |   MinsumTime3SDWhen3RowRight <= -1.799652
|   |   |   |   MinsumTime3SDWhen3RowRight <= -6.288564: NOT (14.0)
|   |   |   |   MinsumTime3SDWhen3RowRight > -6.288564: CONCENTRATING (24.0)
|   |   |   MinsumTime3SDWhen3RowRight > -1.799652: NOT (573.0/104.0)
|   |   MinsumTime3SDWhen3RowRight > 2.049127
|   |   |   MinsumTime3SDWhen3RowRight <= 11.654273: CONCENTRATING (24.0)
|   |   |   MinsumTime3SDWhen3RowRight > 11.654273
|   |   |   |   MinsumTime3SDWhen3RowRight <= 12.159204: NOT (7.0)
|   |   |   |   MinsumTime3SDWhen3RowRight > 12.159204: CONCENTRATING (9.0)
|   SumfrWorkingInSchool > 1
|   |   SumhintTotal <= 10
|   |   |   SumfrWorkingInSchool <= 5
|   |   |   |   SumhintTotal <= 8
|   |   |   |   |   MinsumTime3SDWhen3RowRight <= -5.066348: NOT (8.0/1.0)
|   |   |   |   |   MinsumTime3SDWhen3RowRight > -5.066348
|   |   |   |   |   |   MinsumTime3SDWhen3RowRight <= -1.472802: CONCENTRATING (24.0)
|   |   |   |   |   |   MinsumTime3SDWhen3RowRight > -1.472802
|   |   |   |   |   |   |   MinsumTime3SDWhen3RowRight <= -1.228492: NOT (7.0)
|   |   |   |   |   |   |   MinsumTime3SDWhen3RowRight > -1.228492: CONCENTRATING (104.0/21.0)
|   |   |   |   SumhintTotal > 8: NOT (8.0/1.0)
|   |   |   SumfrWorkingInSchool > 5: NOT (8.0/1.0)
|   |   SumhintTotal > 10: CONCENTRATING (14.0)

Number of Leaves  :     26

Size of the tree :      51
```

Boredom Model for Students from Suburban Schools (REP-Tree)

```
SumfrWorkingInSchool < 0.5
|   MintotalFrTimeOnSkill < 141.88
|   |   MintotalFrTimeOnSkill < 102.45
|   |   |   MintotalFrTimeOnSkill < 66.56
|   |   |   |   MintotalFrTimeOnSkill < 54.74
|   |   |   |   |   SumtotalFrPercentPastWrong < 0.07
|   |   |   |   |   |   MintotalFrTimeOnSkill < 38.36
|   |   |   |   |   |   |   MintotalFrTimeOnSkill < 37.57
|   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 18.01
|   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 17.27 : NOT (422/176) [232/104]
|   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 17.27 : BORED (12/0) [8/0]
|   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 18.01 : NOT (18/0) [8/0]
```

```
|   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 37.57 : BORED (12/0) [8/0]
|   |   |   |   |   |   MintotalFrTimeOnSkill >= 38.36 : NOT (14/0) [6/0]
|   |   |   |   |   SumtotalFrPercentPastWrong >= 0.07 : NOT (79/0) [32/0]
|   |   |   |   MintotalFrTimeOnSkill >= 54.74
|   |   |   |   |   MintotalFrTimeOnSkill < 56.04 : BORED (33/3) [10/0]
|   |   |   |   |   MintotalFrTimeOnSkill >= 56.04
|   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.3 : NOT (10/0) [4/0]
|   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.3
|   |   |   |   |   |   |   MintotalFrTimeOnSkill < 60.73 : NOT (4/0) [1/0]
|   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 60.73
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.42 : BORED (12/0) [8/0]
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.42
|   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.88 : NOT (2/0) [3/0]
|   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.88 : BORED (11/0) [10/1]
|   |   |   MintotalFrTimeOnSkill >= 66.56 : NOT (52/0) [20/0]
|   |   MintotalFrTimeOnSkill >= 102.45
|   |   |   SumtotalFrPercentPastWrong < 0.23
|   |   |   |   MintotalFrTimeOnSkill < 126.23 : NOT (10/0) [5/0]
|   |   |   |   MintotalFrTimeOnSkill >= 126.23
|   |   |   |   |   MintotalFrTimeOnSkill < 129.22 : BORED (17/0) [3/0]
|   |   |   |   |   MintotalFrTimeOnSkill >= 129.22
|   |   |   |   |   |   MintotalFrTimeOnSkill < 136.55 : NOT (5/0) [2/0]
|   |   |   |   |   |   MintotalFrTimeOnSkill >= 136.55
|   |   |   |   |   |   |   MintotalFrTimeOnSkill < 138.46 : BORED (10/0) [10/0]
|   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 138.46 : NOT (4/0) [1/0]
|   |   |   SumtotalFrPercentPastWrong >= 0.23
|   |   |   |   MintotalFrTimeOnSkill < 140.75
|   |   |   |   |   MintotalFrTimeOnSkill < 127.63
|   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.42 : BORED (73/1) [28/0]
|   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.42
|   |   |   |   |   |   |   MintotalFrTimeOnSkill < 125.66
|   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 105.1 : BORED (13/0) [8/1]
|   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 105.1 : NOT (9/0) [9/0]
|   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 125.66 : BORED (14/0) [6/0]
|   |   |   |   |   MintotalFrTimeOnSkill >= 127.63 : NOT (9/0) [6/0]
|   |   |   |   MintotalFrTimeOnSkill >= 140.75 : BORED (67/1) [34/0]
|   MintotalFrTimeOnSkill >= 141.88
|   |   SumtotalFrPercentPastWrong < 0.42
|   |   |   MintotalFrTimeOnSkill < 326.93
|   |   |   |   MintotalFrTimeOnSkill < 163.09
|   |   |   |   |   MintotalFrTimeOnSkill < 162.44 : NOT (14/0) [7/0]
|   |   |   |   |   MintotalFrTimeOnSkill >= 162.44 : BORED (11/0) [10/1]
|   |   |   |   MintotalFrTimeOnSkill >= 163.09 : NOT (65/0) [30/0]
|   |   |   MintotalFrTimeOnSkill >= 326.93
|   |   |   |   MintotalFrTimeOnSkill < 329.25 : BORED (11/0) [9/0]
|   |   |   |   MintotalFrTimeOnSkill >= 329.25
|   |   |   |   |   SumtotalFrPercentPastWrong < 0.27
|   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.16
|   |   |   |   |   |   |   MintotalFrTimeOnSkill < 558.37 : NOT (21/0) [14/0]
|   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 558.37
|   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 567.87 : BORED (13/1) [8/0]
|   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 567.87 : NOT (13/0) [1/0]
|   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.16
|   |   |   |   |   |   |   MintotalFrTimeOnSkill < 1036.04
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.17
|   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 652.43 : NOT (2/0) [1/0]
|   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 652.43 : BORED (17/0) [5/2]
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.17
|   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 407.74
|   |   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 374.26 : NOT (4/0) [3/0]
|   |   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 374.26 : BORED (14/0) [6/0]
|   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 407.74 : NOT (14/0) [9/0]
|   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 1036.04 : BORED (14/0) [6/0]
|   |   |   |   |   SumtotalFrPercentPastWrong >= 0.27 : NOT (29/0) [8/0]
|   |   SumtotalFrPercentPastWrong >= 0.42 : NOT (83/0) [47/0]
SumfrWorkingInSchool >= 0.5
|   SumfrWorkingInSchool < 1.5
|   |   MintotalFrTimeOnSkill < 223.6
|   |   |   MintotalFrTimeOnSkill < 212.38
|   |   |   |   MintotalFrTimeOnSkill < 164.74
|   |   |   |   |   MintotalFrTimeOnSkill < 116.82
|   |   |   |   |   |   MintotalFrTimeOnSkill < 34.1
|   |   |   |   |   |   |   MintotalFrTimeOnSkill < 28.73
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.58
|   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 23.09
|   |   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 22.21
|   |   |   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 4.82
|   |   |   |   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill < 0.01
|   |   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.36
|   |   |   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.1 : BORED (246/49) [105/22]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.1 : NOT (2/0) [1/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.36 : BORED (17/2) [6/1]
|   |   |   |   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 0.01 : BORED (15/0) [5/0]
|   |   |   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 4.82 : NOT (7/0) [3/0]
|   |   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 22.21 : BORED (24/0) [16/0]
|   |   |   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 23.09 : NOT (5/0) [0/0]
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.58 : NOT (5/0) [4/0]
|   |   |   |   |   |   |   MintotalFrTimeOnSkill >= 28.73 : BORED (59/1) [25/3]
|   |   |   |   |   |   MintotalFrTimeOnSkill >= 34.1 : NOT (17/0) [19/0]
|   |   |   |   |   MintotalFrTimeOnSkill >= 116.82
|   |   |   |   |   |   MintotalFrTimeOnSkill < 130.69 : BORED (27/0) [14/1]
|   |   |   |   |   |   MintotalFrTimeOnSkill >= 130.69
|   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.13 : NOT (3/0) [3/0]
|   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.13 : BORED (41/3) [28/6]
|   |   |   |   MintotalFrTimeOnSkill >= 164.74 : NOT (13/0) [4/0]
```

```
|   |   |   MintotalFrTimeOnSkill >= 212.38 : BORED (39/0) [21/0]
|   |   MintotalFrTimeOnSkill >= 223.6
|   |   |   MintotalFrTimeOnSkill < 552.65 : NOT (33/0) [16/0]
|   |   |   MintotalFrTimeOnSkill >= 552.65
|   |   |   |   SumtotalFrPercentPastWrong < 0.19 : NOT (4/0) [2/0]
|   |   |   |   SumtotalFrPercentPastWrong >= 0.19
|   |   |   |   |   MintotalFrTimeOnSkill < 761.93 : BORED (36/2) [28/2]
|   |   |   |   |   MintotalFrTimeOnSkill >= 761.93 : NOT (2/0) [4/0]
|   SumfrWorkingInSchool >= 1.5
|   |   MintotalFrTimeOnSkill < 10.25
|   |   |   MintotalFrTimeOnSkill < 5.09
|   |   |   |   SumfrWorkingInSchool < 5 : NOT (57/0) [24/0]
|   |   |   |   SumfrWorkingInSchool >= 5 : BORED (14/1) [10/3]
|   |   |   MintotalFrTimeOnSkill >= 5.09 : BORED (16/0) [4/0]
|   |   MintotalFrTimeOnSkill >= 10.25 : NOT (63/0) [24/0]

Size of the tree : 123
```

Confusion Model for Students from Suburban Schools (REP-Tree)

```
MaxtotalFrAttempted < 112.5
|   AveragetotalFrAttempted < 5.75
|   |   MinresponseIsFillIn < 0.5 : NOT (14/0) [8/0]
|   |   MinresponseIsFillIn >= 0.5
|   |   |   MaxtotalFrAttempted < 1.5 : NOT (6/0) [7/0]
|   |   |   MaxtotalFrAttempted >= 1.5
|   |   |   |   Maxcorrect < 0.5 : CONFUSED (449/6) [223/6]
|   |   |   |   Maxcorrect >= 0.5
|   |   |   |   |   MaxtotalFrAttempted < 2.5 : CONFUSED (144/3) [79/0]
|   |   |   |   |   MaxtotalFrAttempted >= 2.5 : NOT (12/0) [9/0]
|   AveragetotalFrAttempted >= 5.75
|   |   SumtotalFrPercentPastWrong < 0.17
|   |   |   MaxtotalFrSkillOpportunitiesByScaffolding < 0.47 : NOT (238/0) [132/0]
|   |   |   MaxtotalFrSkillOpportunitiesByScaffolding >= 0.47
|   |   |   |   MaxtotalFrSkillOpportunitiesByScaffolding < 0.97 : CONFUSED (67/0) [45/2]
|   |   |   |   MaxtotalFrSkillOpportunitiesByScaffolding >= 0.97 : NOT (30/0) [11/0]
|   |   SumtotalFrPercentPastWrong >= 0.17
|   |   |   SumtotalFrPercentPastWrong < 0.85
|   |   |   |   MaxtotalFrAttempted < 14.5 : NOT (31/0) [9/0]
|   |   |   |   MaxtotalFrAttempted >= 14.5
|   |   |   |   |   SumbottomHint < 0.5
|   |   |   |   |   |   MaxtotalFrAttempted < 74
|   |   |   |   |   |   |   Mincorrect < 0.5 : NOT (16/0) [10/0]
|   |   |   |   |   |   |   Mincorrect >= 0.5
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.75
|   |   |   |   |   |   |   |   |   MaxtotalFrAttempted < 21.5
|   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.48
|   |   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.19 : CONFUSED (72/0) [38/0]
|   |   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.19
|   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong < 0.43 : NOT (7/0) [2/0]
|   |   |   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.43 : CONFUSED (70/0) [40/0]
|   |   |   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.48 : NOT (5/0) [0/0]
|   |   |   |   |   |   |   |   |   MaxtotalFrAttempted >= 21.5 : NOT (16/0) [9/0]
|   |   |   |   |   |   |   |   SumtotalFrPercentPastWrong >= 0.75 : CONFUSED (78/0) [32/0]
|   |   |   |   |   |   MaxtotalFrAttempted >= 74 : NOT (27/0) [16/0]
|   |   |   |   |   SumbottomHint >= 0.5 : CONFUSED (156/1) [67/2]
|   |   |   SumtotalFrPercentPastWrong >= 0.85 : NOT (78/0) [33/0]
MaxtotalFrAttempted >= 112.5 : NOT (534/0) [255/0]

Size of the tree : 39
```

Frustration Model for Students from Suburban Schools (REP-Tree)

```
Sumscaffold < 0.5
|   AveragetimeTaken < 4.49
|   |   AveragehelpAccessUnder2Sec < 0.37
|   |   |   AveragefrIsHelpRequest < 0.31
|   |   |   |   AveragetimeTaken < 4.45
|   |   |   |   |   AveragetimeTaken < 2.01 : FRUSTRATED (61/0) [36/1]
|   |   |   |   |   AveragetimeTaken >= 2.01 : NOT (29/0) [14/0]
|   |   |   |   AveragetimeTaken >= 4.45 : FRUSTRATED (70/0) [26/0]
|   |   |   AveragefrIsHelpRequest >= 0.31 : FRUSTRATED (147/1) [48/2]
|   |   AveragehelpAccessUnder2Sec >= 0.37 : NOT (14/0) [3/0]
|   AveragetimeTaken >= 4.49
|   |   AveragetimeTaken < 10.12 : NOT (83/0) [40/0]
|   |   AveragetimeTaken >= 10.12
|   |   |   AveragetimeTaken < 23.13
|   |   |   |   AveragetimeTaken < 22.94
|   |   |   |   |   AveragetimeTaken < 17.82
|   |   |   |   |   |   AveragetimeTaken < 10.15 : FRUSTRATED (63/0) [33/0]
|   |   |   |   |   |   AveragetimeTaken >= 10.15
|   |   |   |   |   |   |   AveragetimeTaken < 17.72
|   |   |   |   |   |   |   |   AveragetimeTaken < 11.93 : NOT (27/0) [16/0]
|   |   |   |   |   |   |   |   AveragetimeTaken >= 11.93
|   |   |   |   |   |   |   |   |   AveragetimeTaken < 11.97 : FRUSTRATED (63/0) [34/1]
|   |   |   |   |   |   |   |   |   AveragetimeTaken >= 11.97
|   |   |   |   |   |   |   |   |   |   AveragetimeTaken < 16.03 : NOT (76/0) [33/0]
|   |   |   |   |   |   |   |   |   |   AveragetimeTaken >= 16.03
```

```
|   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken < 16.1 : FRUSTRATED (63/0) [35/2]
|   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken >= 16.1 : NOT (24/0) [11/0]
|   |   |   |   |   |   |   AveragetimeTaken >= 17.72 : FRUSTRATED (62/0) [38/4]
|   |   |   |   |   AveragetimeTaken >= 17.82 : NOT (72/0) [38/0]
|   |   |   |   AveragetimeTaken >= 22.94 : FRUSTRATED (123/0) [69/0]
|   |   |   AveragetimeTaken >= 23.13
|   |   |   |   AveragetimeTaken < 49.28 : NOT (230/0) [106/0]
|   |   |   |   AveragetimeTaken >= 49.28
|   |   |   |   |   AveragetimeTaken < 49.46 : FRUSTRATED (68/0) [28/0]
|   |   |   |   |   AveragetimeTaken >= 49.46
|   |   |   |   |   |   AveragetimeTaken < 252.17
|   |   |   |   |   |   |   AveragetimeTaken < 249.32
|   |   |   |   |   |   |   |   AveragetimeTaken < 58.23 : NOT (42/0) [30/0]
|   |   |   |   |   |   |   |   AveragetimeTaken >= 58.23
|   |   |   |   |   |   |   |   |   AveragetimeTaken < 58.28 : FRUSTRATED (63/0) [33/0]
|   |   |   |   |   |   |   |   |   AveragetimeTaken >= 58.28
|   |   |   |   |   |   |   |   |   |   AveragetimeTaken < 75.7 : NOT (71/0) [34/0]
|   |   |   |   |   |   |   |   |   |   AveragetimeTaken >= 75.7
|   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken < 76.2 : FRUSTRATED (64/0) [34/2]
|   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken >= 76.2
|   |   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken < 106.13 : NOT (70/0) [36/0]
|   |   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken >= 106.13
|   |   |   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken < 106.57 : FRUSTRATED (59/0) [37/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken >= 106.57
|   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken < 173.79 : NOT (76/0) [39/0]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken >= 173.79
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken < 175.26 : FRUSTRATED (60/0) [37/1]
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   AveragetimeTaken >= 175.26 : NOT (36/0) [22/0]
|   |   |   |   |   |   |   AveragetimeTaken >= 249.32 : FRUSTRATED (59/0) [37/0]
|   |   |   |   |   |   AveragetimeTaken >= 252.17 : NOT (48/0) [23/0]
Sumscaffold >= 0.5 : NOT (123/0) [53/0]

Size of the tree : 57
```

APPENDIX B

High School Survey Questions

(Questions procured from ASSISTments Team)

1. What is the name of your high school? What year are you in?
   Example: Shrewsbury High School, Senior Year
   ____

2. What mathematics course are you taking right now? If you are not taking a mathematics course now, what is the last (most recent) mathematics course you took?
   ____

3. What science course are you taking right now? If you are not taking a science course now, what is the last (most recent) science course you took?
   ____

4. What current or past jobs have you taken while in High School?
   Examples: Part-time Tutor, Lab Assistant, Volunteer, Cashier, Restaurant Staff, None
   ____

5. What are you planning to do after high school? (Check all that apply)
   Examples: College, Work, College+Work, None of the Above
   ☐College (Includes Community College, Trade and Technical Schools, Online Education, etc.)
   ☐Work (Includes Full-time, Part-time, Internship, Volunteering, etc.)

6. If you plan to go to college after high school, what major or majors do you find most interesting?
   Example: Pre-Medicine, Business, Engineering, Film
   ____

7. If you plan to work after high school (whether or not you go to college), what kind of jobs are you interested in?
   ____

8. Please write your email address, and if possible your Facebook screen name/webpage so that we can contact you in the future, to participate in future research.
   ____

APPENDIX C

Post-High School Survey Questions

(Questions procured from ASSISTments Team)

* Answer Required

1. What is your name?

____

2. What is the name of your high school? What year did you graduate?
   Example: Shrewsbury High School, 2010

____

3. Are you currently attending college, a technical institute, a trade school, distance education, or any other educational program? If the answer is NO, please proceed to question #4. *
   - ☐ YES
   - ☐ NO

4. Please name the institution (college, school) you are attending. *
   Examples: Worcester Polytechnic Institute, Harvard University, University of Phoenix, Quinsigamond Community College, Porter and Chester Institute

____

5. Please name the major, program, or training course you are taking. If you do not have a major, list the major you are most likely to complete. It is acceptable to list multiple majors. *
   Examples: Physics, English Literature, Automotive Technology, Practical Nursing

____

6. If you are currently employed, either part-time or full-time, please list your current job/ job title. *
   Examples: Technician, Research Assistant, Teacher, Salesperson

____

7. Please list your current area of work. *
   Examples: Law, Product Development, Food Service, Retail, Medical

____

8. What was the last MATH course you took in High School? *
   Examples: Algebra, Geometry, Trigonometry, Statistics, Finance

____

9. Would it be OK to contact you again? *
   - ☐ YES
   - ☐ NO

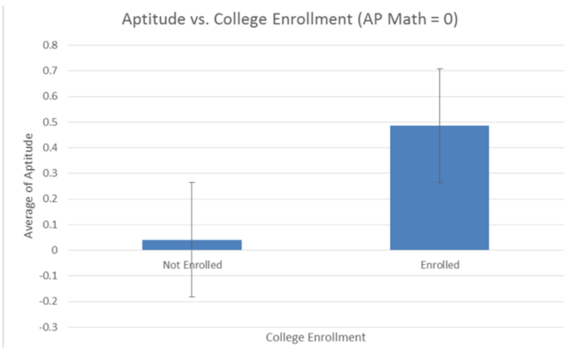
10. So that we can send you your gift certificate, please write your email address. *

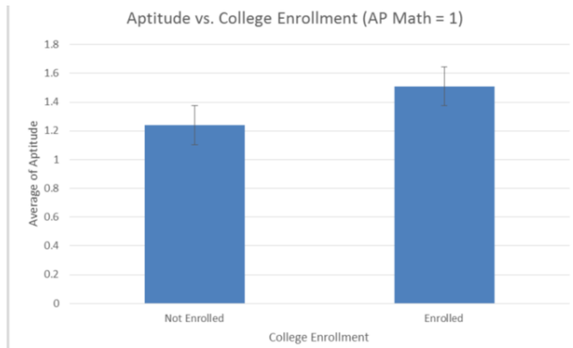
____

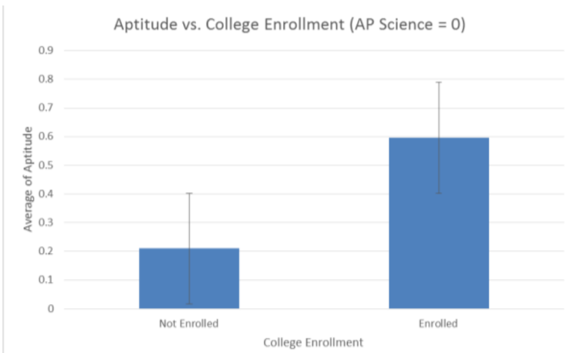
APPENDIX D

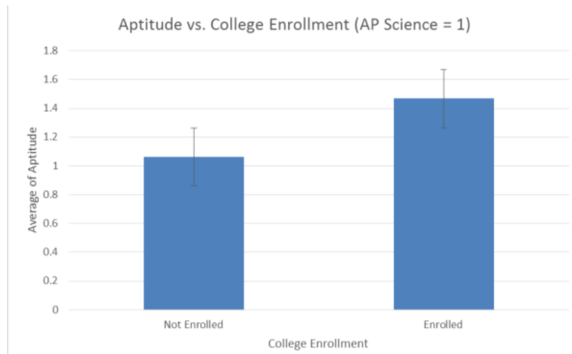
Graphs of Middle School Performance-Engagement Factors vs. College Attendance Outcomes in Study 3

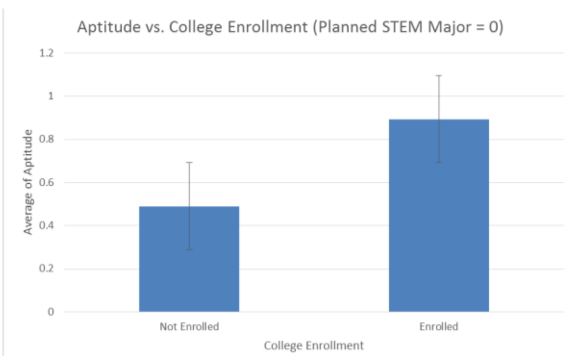
Aptitude vs. College Enrollment

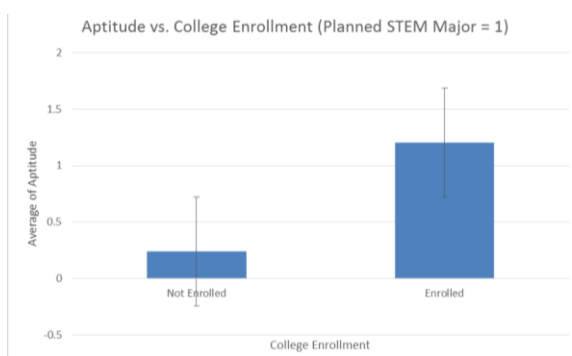


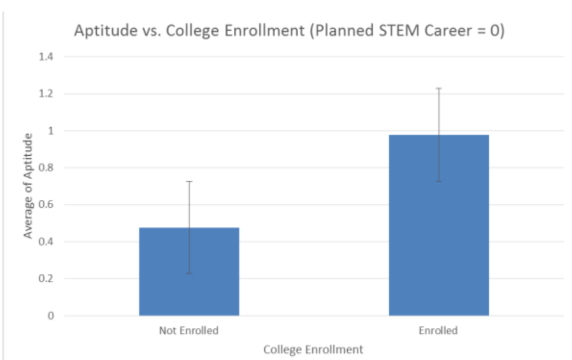


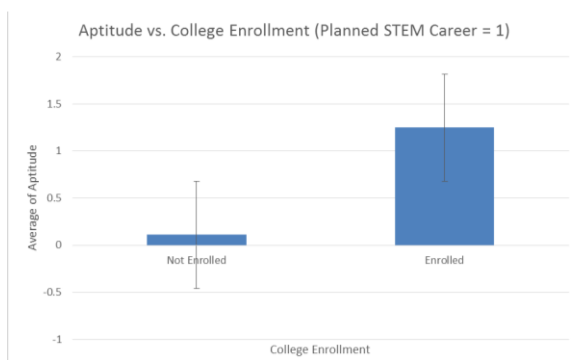




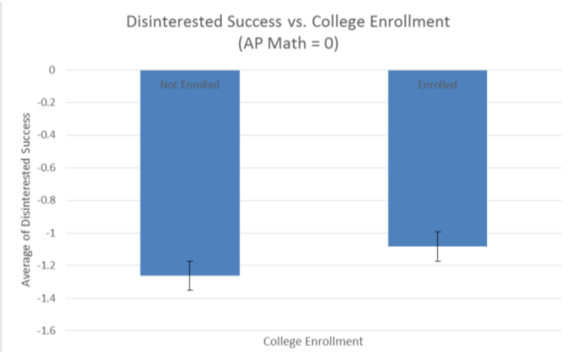


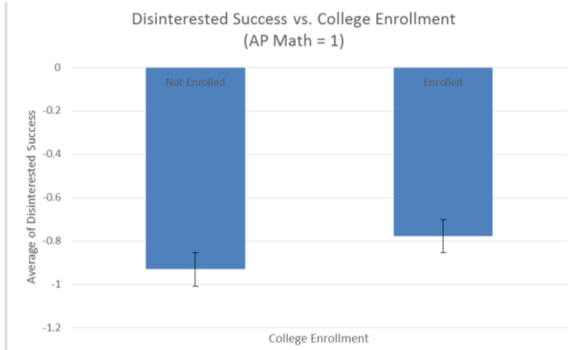


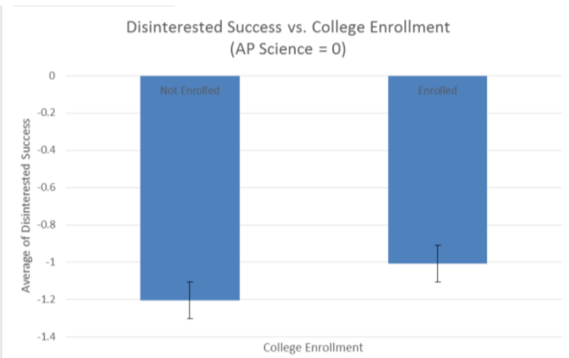


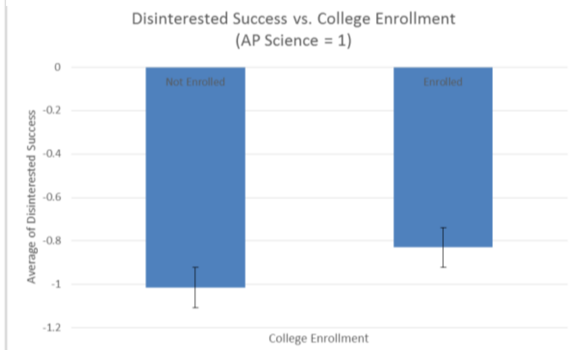


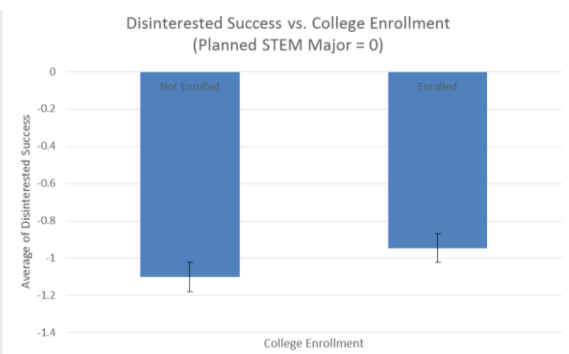
Disinterested Success vs. College Enrollment

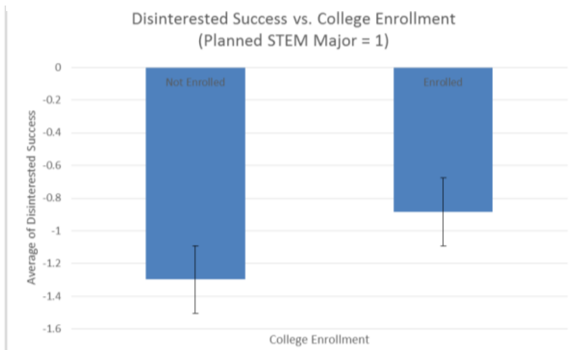


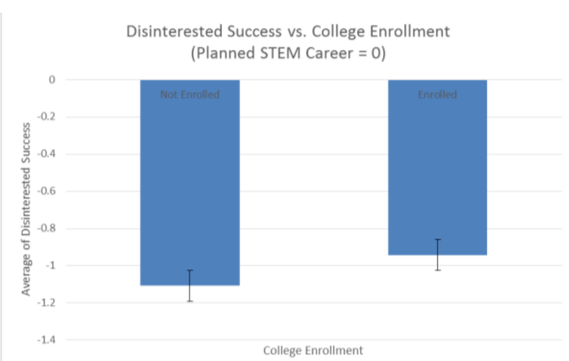


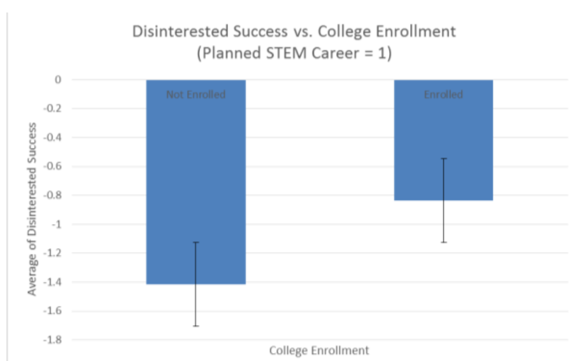




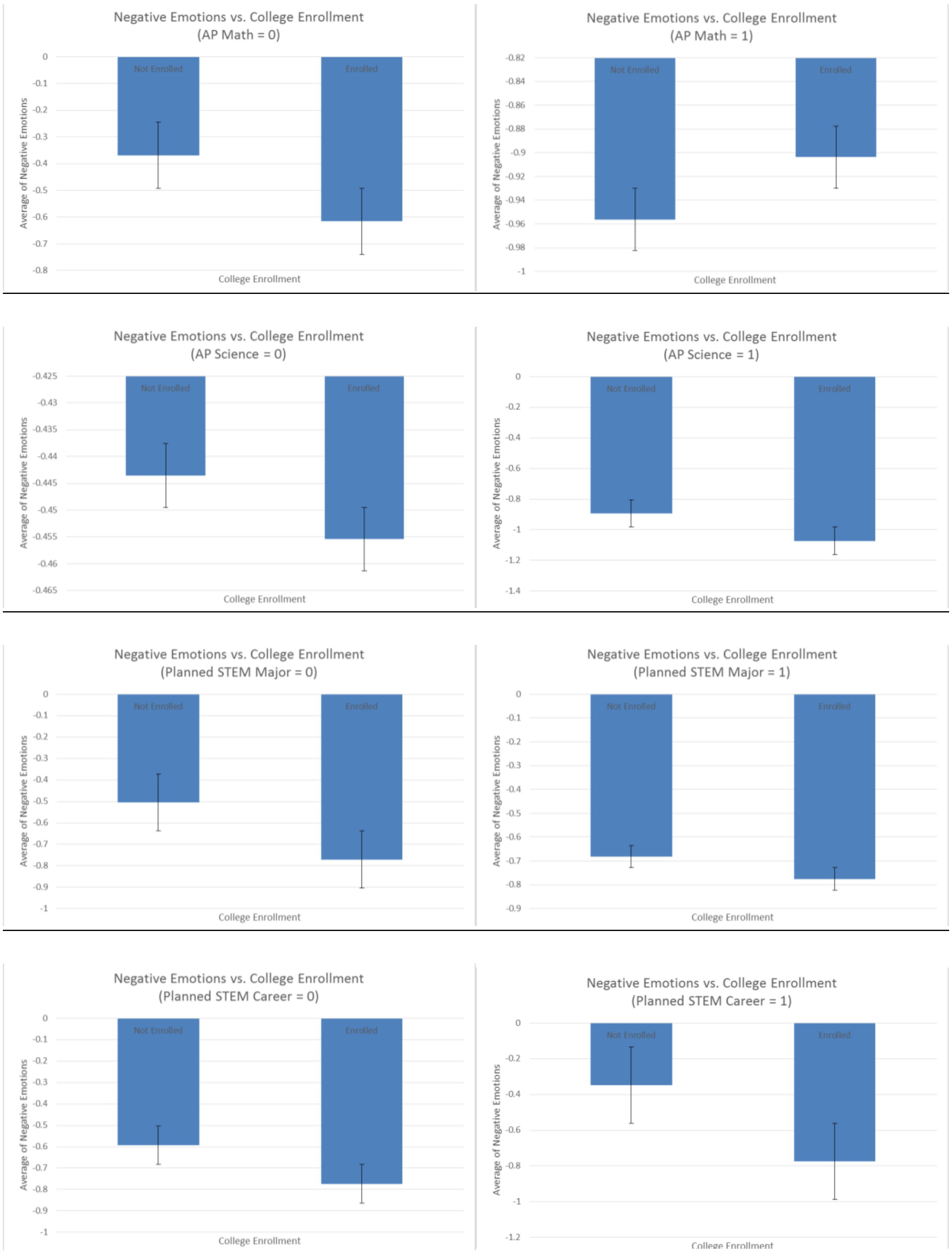




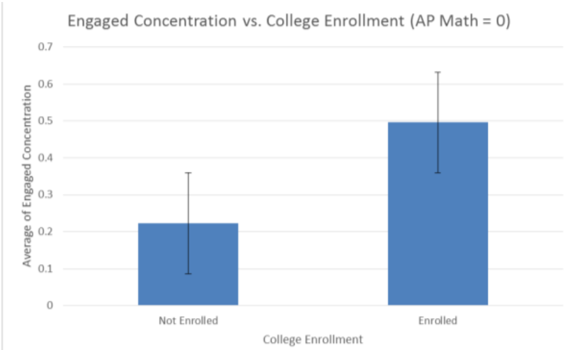


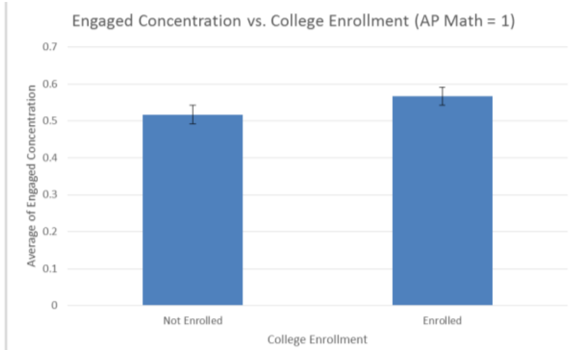


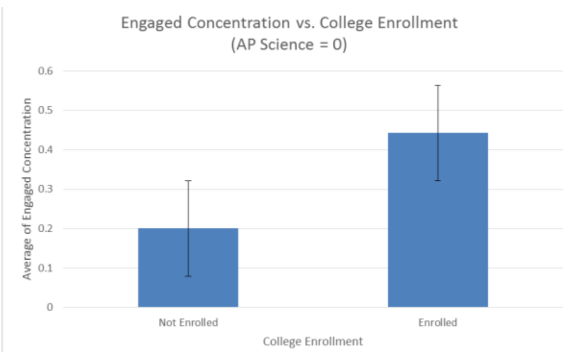
Negative Emotions vs. College Enrollment

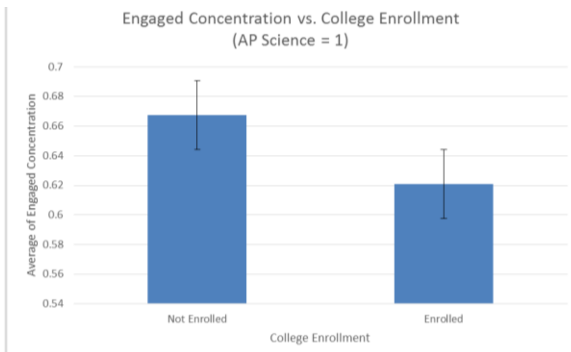


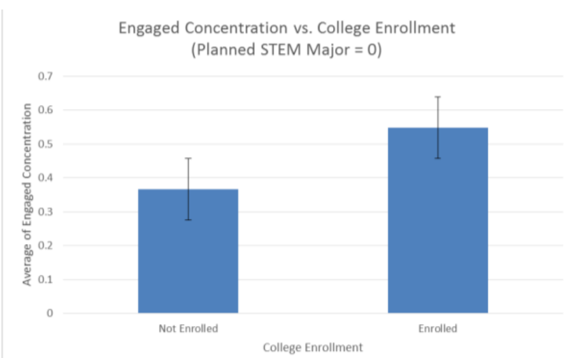
Engaged Concentration vs. College Enrollment

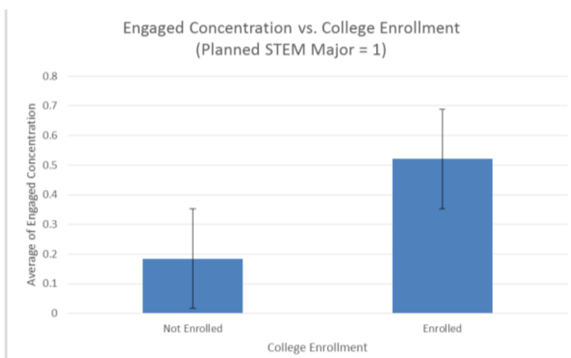


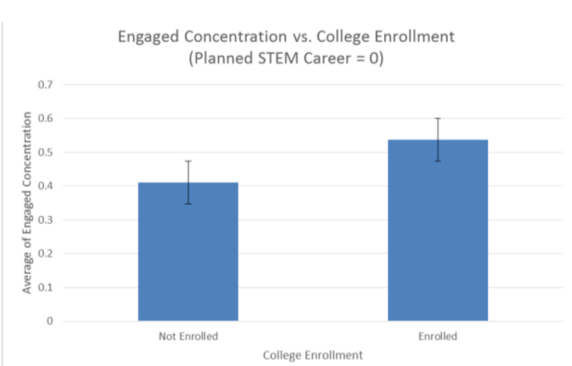


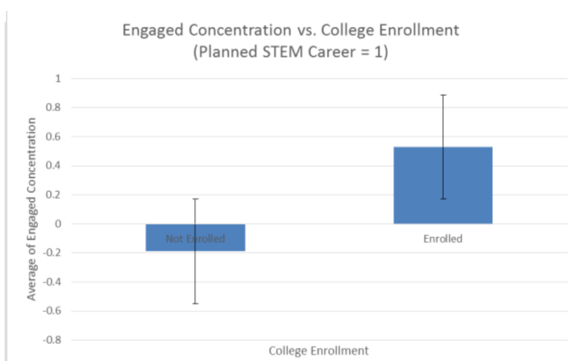




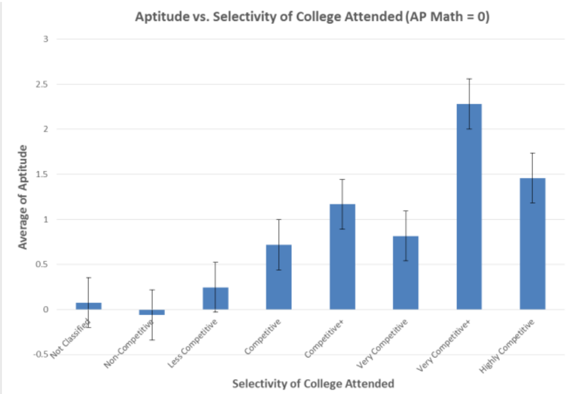


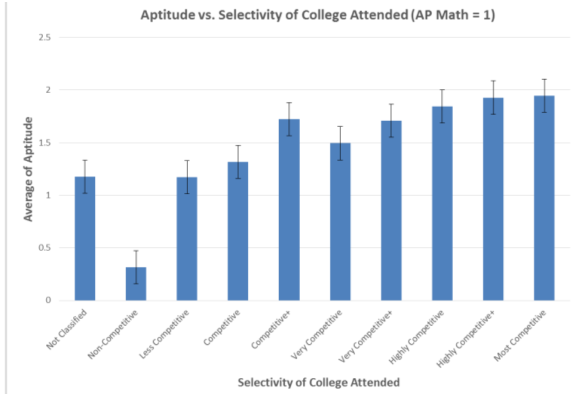


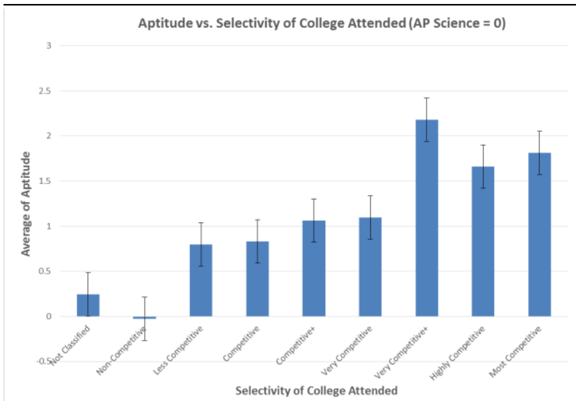


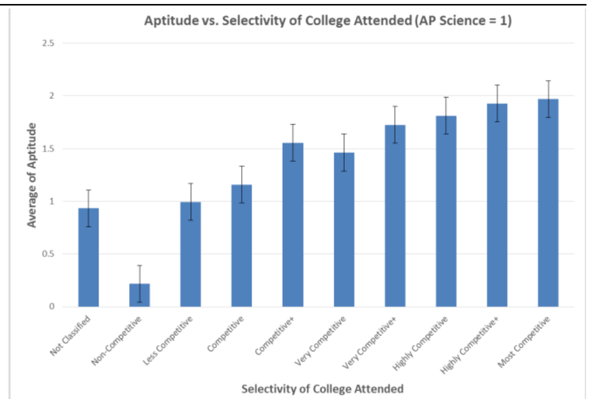


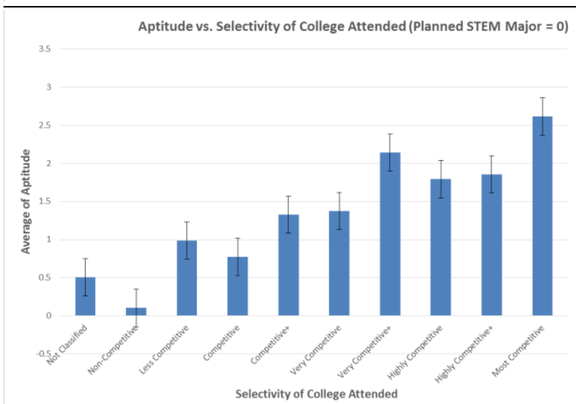
Aptitude vs. Selectivity of College Attended

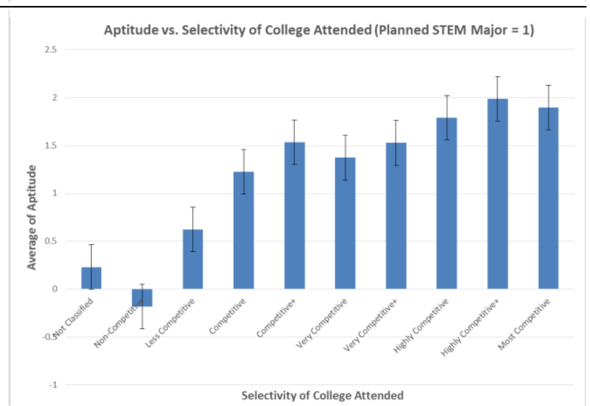


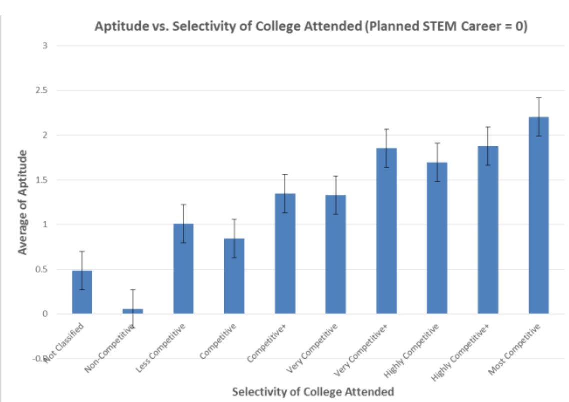


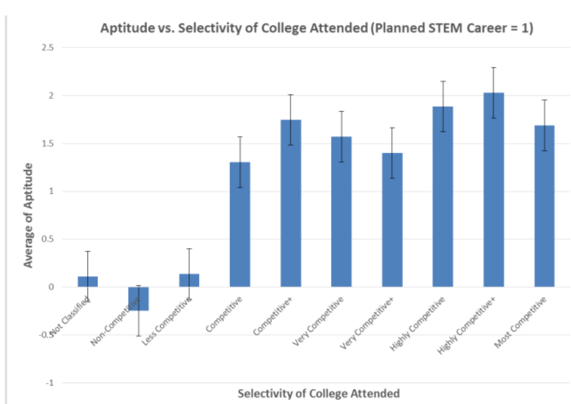




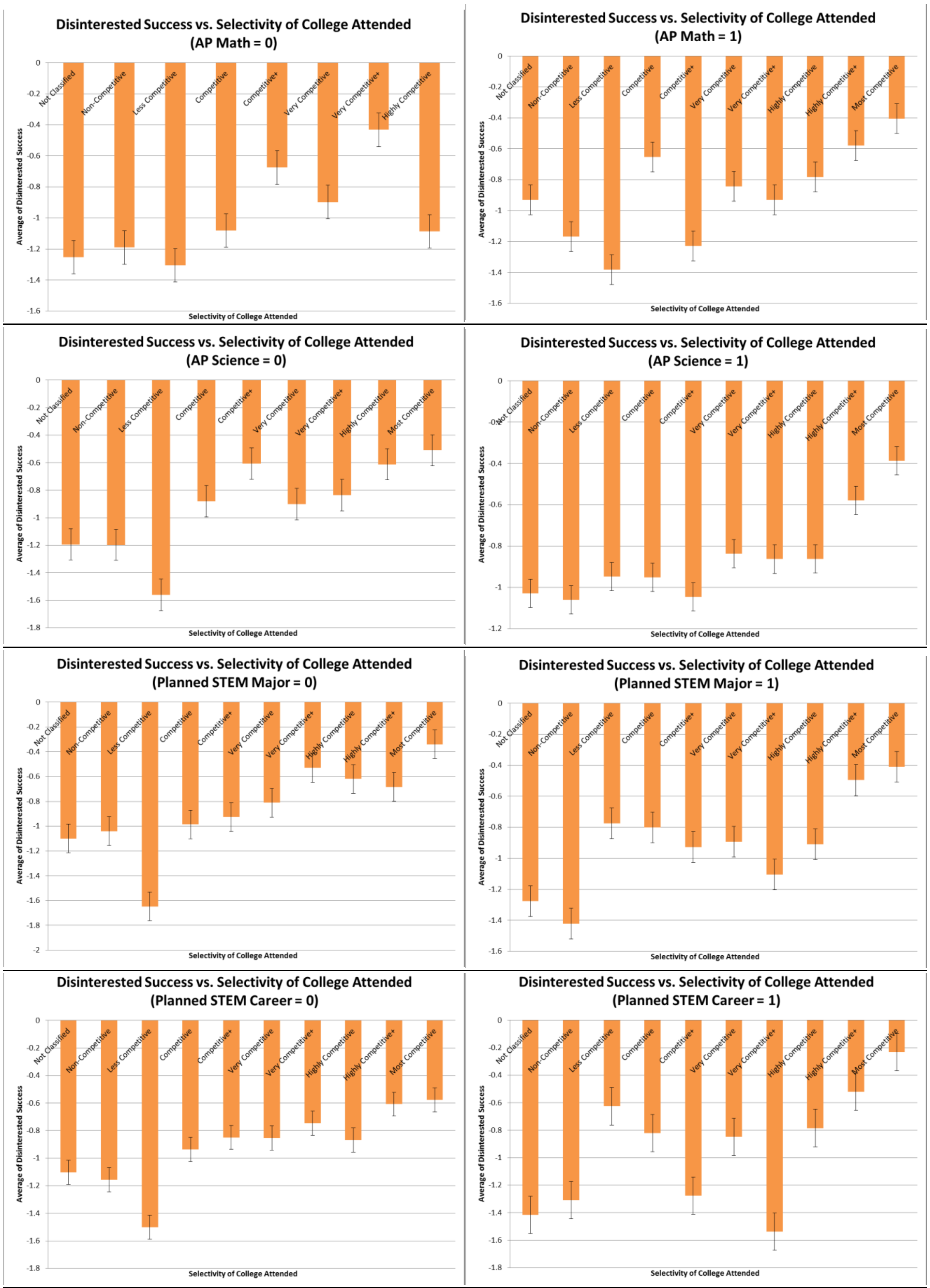




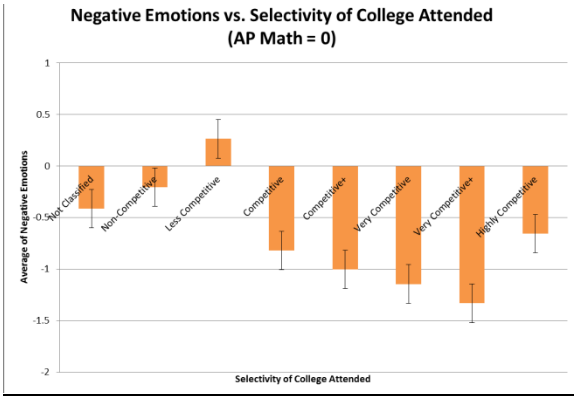


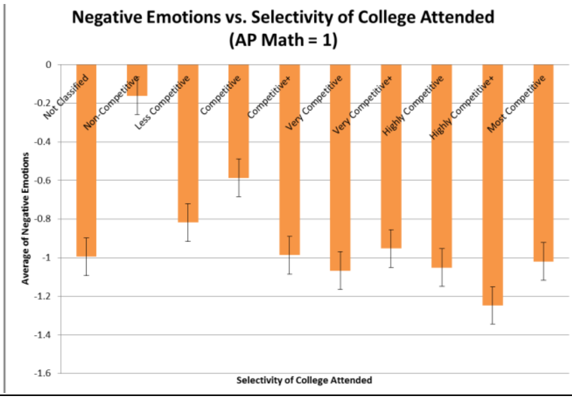


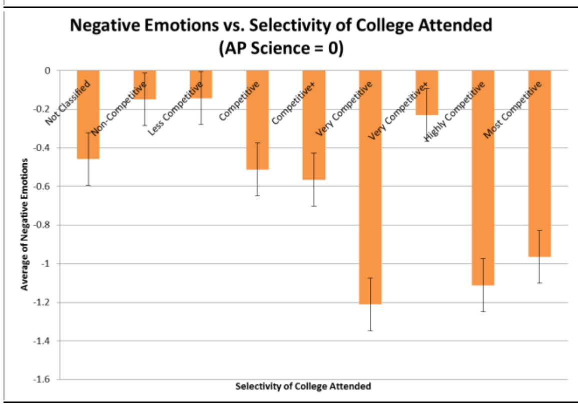
Disinterested Success vs. Selectivity of College Attended

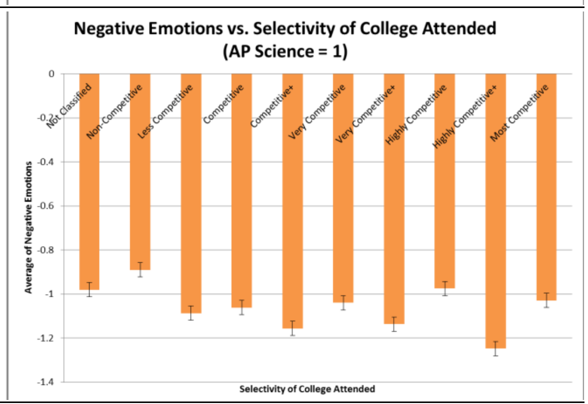


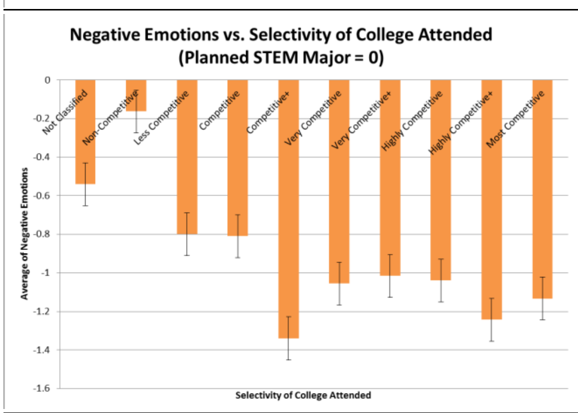
<u>Negative Emotions vs. Selectivity of College Attended</u>

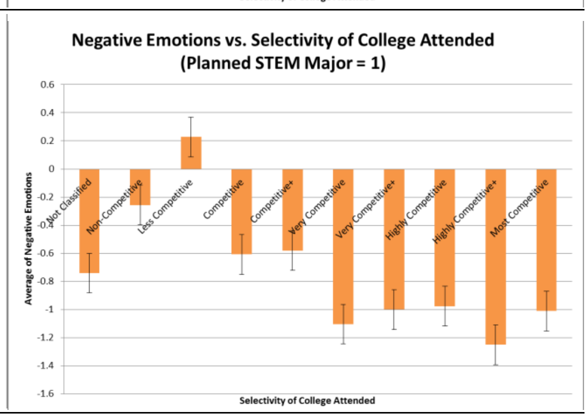


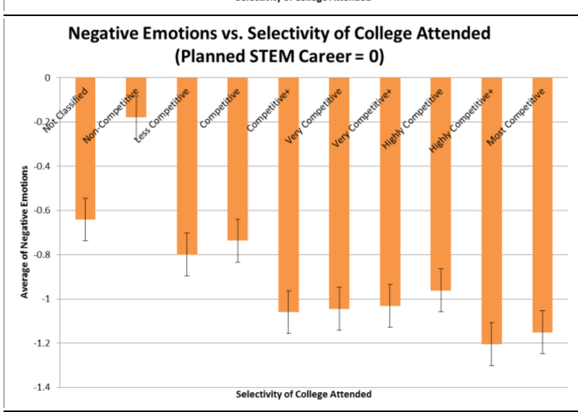


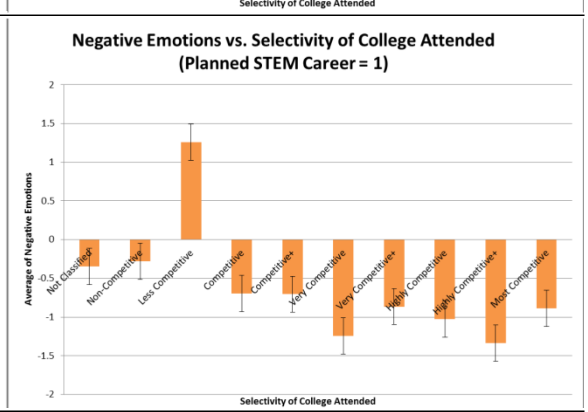




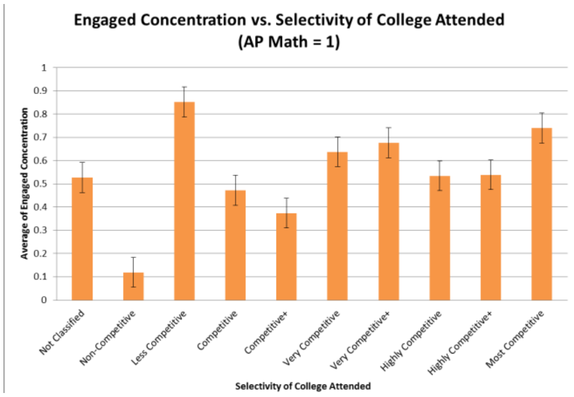


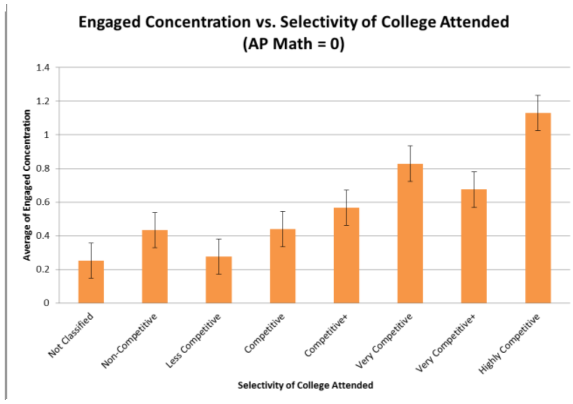


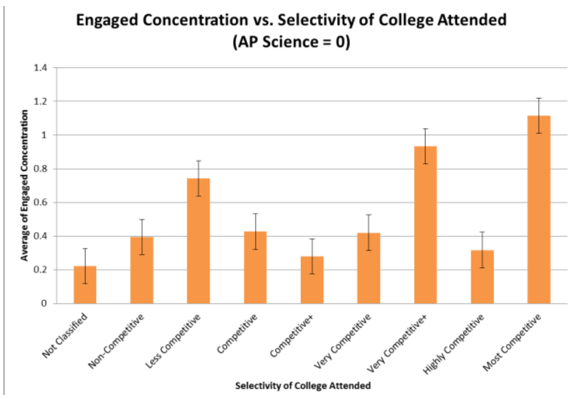


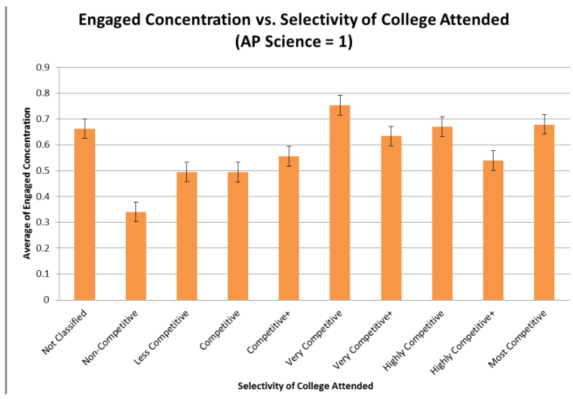


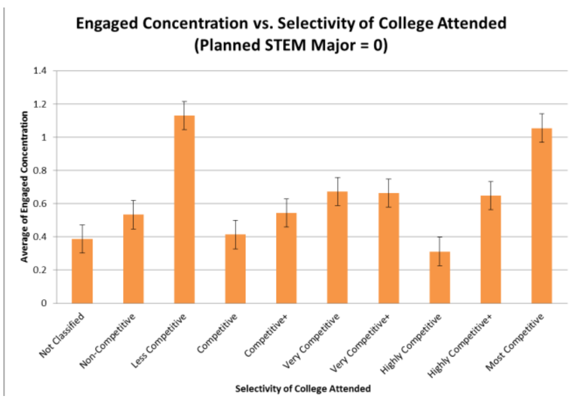
<u>*Engaged Concentration vs. Selectivity of College Attended*</u>

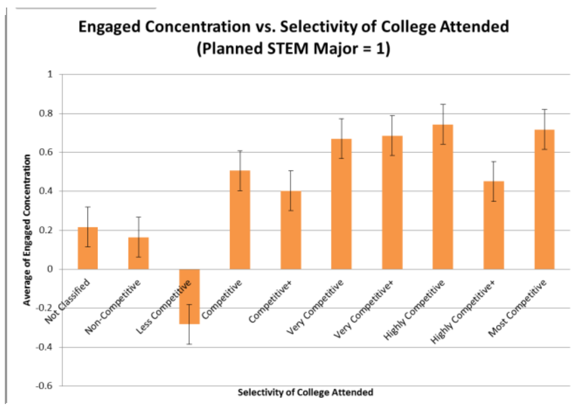


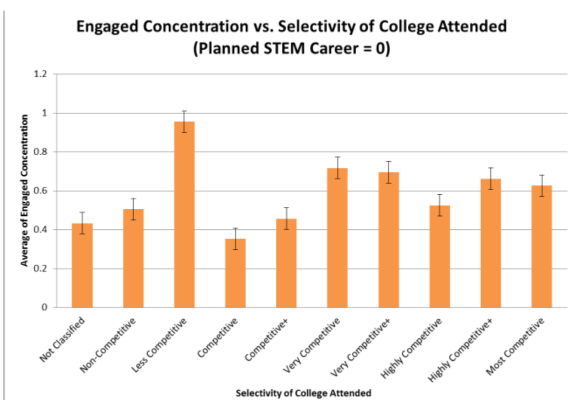


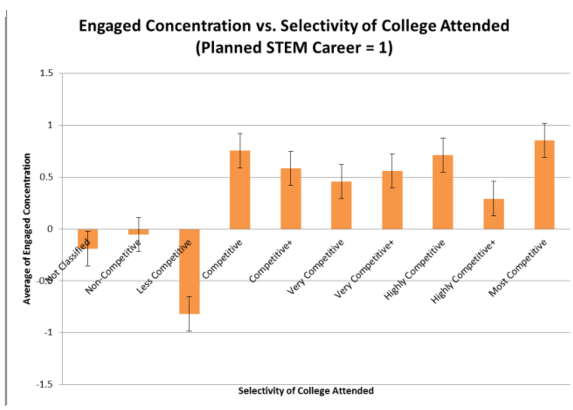












APPENDIX E

MPlus Outputs of Modeling College Enrollment from Middle School Performance-Engagement Factor with Mediation from High School Variables in Study 3

1. High school mediator treated as continuous variable (From Table 27)

<u>Aptitude → AP Math → College Enrollment.</u>

```
MODEL FIT INFORMATION

Number of Free Parameters                        6

Loglikelihood

          H0 Value                        -281.370
          H0 Scaling Correction Factor      0.8658
            for MLR

Information Criteria

          Akaike (AIC)                     574.740
          Bayesian (BIC)                   596.592
          Sample-Size Adjusted BIC         577.566
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL1                0.595      0.188      3.169      0.002
    MATH               0.322      0.365      0.883      0.377

 MATH       ON
    FL1                0.292      0.020     14.953      0.000

 Intercepts
    MATH               0.232      0.030      7.735      0.000

 Thresholds
    ENROLL$1          -0.800      0.196     -4.074      0.000

 Residual Variances
    MATH               0.169      0.011     14.839      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 ENROLL     ON
    FL1                1.813
    MATH               1.380
```

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ENROLL ON				
FL1	0.299	0.089	3.365	0.001
MATH	0.083	0.093	0.890	0.374
MATH ON				
FL1	0.570	0.041	13.867	0.000
Intercepts				
MATH	0.463	0.061	7.602	0.000
Thresholds				
ENROLL$1	-0.413	0.109	-3.791	0.000
Residual Variances				
MATH	0.675	0.047	14.417	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ENROLL	0.125	0.053	2.343	0.019
MATH	0.325	0.047	6.934	0.000

<u>Aptitude → AP Science → College Enrollment.</u>

MODEL FIT INFORMATION

Number of Free Parameters	6
Loglikelihood	
H0 Value	-297.559
H0 Scaling Correction Factor for MLR	0.8567
Information Criteria	
Akaike (AIC)	607.118
Bayesian (BIC)	628.969
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	609.943

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL1                0.529      0.184      2.867      0.004
    SCIENCE            0.714      0.376      1.900      0.057

 SCIENCE    ON
    FL1                0.245      0.022     11.262      0.000

 Intercepts
    SCIENCE            0.236      0.031      7.624      0.000

 Thresholds
    ENROLL$1          -0.721      0.193     -3.736      0.000

 Residual Variances
    SCIENCE            0.191      0.011     17.857      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 ENROLL     ON
    FL1                1.697
    SCIENCE            2.042


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL1                0.263      0.087      3.010      0.003
    SCIENCE            0.181      0.093      1.956      0.051

 SCIENCE    ON
    FL1                0.479      0.045     10.603      0.000

 Intercepts
    SCIENCE            0.473      0.061      7.749      0.000

 Thresholds
    ENROLL$1          -0.367      0.106     -3.467      0.001

 Residual Variances
    SCIENCE            0.770      0.043     17.790      0.000
```

```
R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    ENROLL             0.147      0.057      2.587      0.010
    SCIENCE            0.230      0.043      5.302      0.000
```

Aptitude → Planned STEM Major → College Enrollment.

```
MODEL FIT INFORMATION

Number of Free Parameters                        6

Loglikelihood

          H0 Value                        -329.471
          H0 Scaling Correction Factor      0.8471
            for MLR

Information Criteria

          Akaike (AIC)                     670.942
          Bayesian (BIC)                   692.793
          Sample-Size Adjusted BIC         673.767
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL1                0.649      0.170      3.818      0.000
    STEMMAJ            0.765      0.333      2.296      0.022

 STEMMAJ    ON
    FL1                0.077      0.030      2.586      0.010

 Intercepts
    STEMMAJ            0.376      0.040      9.492      0.000

 Thresholds
    ENROLL$1          -0.609      0.206     -2.963      0.003

 Residual Variances
    STEMMAJ            0.242      0.005     45.534      0.000
```

```
LOGISTIC REGRESSION ODDS RATIO RESULTS

 ENROLL     ON
    FL1                1.913
    STEMMAJ            2.149


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL1                0.320      0.077      4.152      0.000
    STEMMAJ            0.192      0.081      2.364      0.018

 STEMMAJ    ON
    FL1                0.151      0.059      2.572      0.010

 Intercepts
    STEMMAJ            0.757      0.076      9.904      0.000

 Thresholds
    ENROLL$1          -0.308      0.110     -2.794      0.005

 Residual Variances
    STEMMAJ            0.977      0.018     55.150      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    ENROLL             0.158      0.054      2.905      0.004
    STEMMAJ            0.023      0.018      1.286      0.198
```

<u>Aptitude → Planned STEM Career → College Enrollment.</u>

```
MODEL FIT INFORMATION

Number of Free Parameters                        6

Loglikelihood

          H0 Value                        -286.636
          H0 Scaling Correction Factor      0.9910
            for MLR
```

```
Information Criteria

          Akaike (AIC)                     585.273
          Bayesian (BIC)                   607.124
          Sample-Size Adjusted BIC         588.098
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL1                0.679      0.168      4.051      0.000
    STEMCAR            0.488      0.400      1.222      0.222

 STEMCAR    ON
    FL1                0.042      0.027      1.525      0.127

 Intercepts
    STEMCAR            0.192      0.034      5.731      0.000

 Thresholds
    ENROLL$1          -0.781      0.191     -4.097      0.000

 Residual Variances
    STEMCAR            0.176      0.013     13.069      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 ENROLL     ON
    FL1                1.971
    STEMCAR            1.629


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL1                0.340      0.076      4.492      0.000
    STEMCAR            0.106      0.086      1.231      0.218

 STEMCAR    ON
    FL1                0.096      0.063      1.529      0.126

 Intercepts
    STEMCAR            0.457      0.069      6.574      0.000

 Thresholds
```

```
   ENROLL$1          -0.401      0.105     -3.814      0.000

 Residual Variances
   STEMCAR            0.991      0.012     81.833      0.000


R-SQUARE

   Observed                                        Two-Tailed
   Variable        Estimate       S.E.  Est./S.E.    P-Value

   ENROLL             0.134      0.053      2.514      0.012
   STEMCAR            0.009      0.012      0.764      0.445
```

<u>Disinterested Success → AP Math → College Enrollment.</u>

```
MODEL FIT INFORMATION

Number of Free Parameters                        6

Loglikelihood

          H0 Value                        -332.780
          H0 Scaling Correction Factor      0.8729
            for MLR

Information Criteria

          Akaike (AIC)                     677.560
          Bayesian (BIC)                   699.412
          Sample-Size Adjusted BIC         680.386
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                   Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
   FL2                0.342      0.226      1.516      0.129
   MATH               0.850      0.325      2.616      0.009

 MATH       ON
   FL2                0.158      0.041      3.812      0.000

 Intercepts
   MATH               0.652      0.051     12.738      0.000

 Thresholds
   ENROLL$1          -1.363      0.335     -4.072      0.000
```

```
Residual Variances
    MATH               0.237      0.007     35.655      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 ENROLL     ON
    FL2                1.408
    MATH               2.339


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL2                0.132      0.086      1.538      0.124
    MATH               0.225      0.082      2.744      0.006

 MATH       ON
    FL2                0.230      0.061      3.780      0.000

 Intercepts
    MATH               1.304      0.101     12.866      0.000

 Thresholds
    ENROLL$1          -0.720      0.176     -4.084      0.000

 Residual Variances
    MATH               0.947      0.028     33.941      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    ENROLL             0.081      0.044      1.838      0.066
    MATH               0.053      0.028      1.890      0.059
```

Disinterested Success → AP Science → College Enrollment.

```
MODEL FIT INFORMATION

Number of Free Parameters                        6

Loglikelihood

          H0 Value                        -334.308
          H0 Scaling Correction Factor      0.8586
            for MLR

Information Criteria

          Akaike (AIC)                     680.616
          Bayesian (BIC)                   702.468
          Sample-Size Adjusted BIC         683.442
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL2                0.352      0.212      1.660      0.097
    SCIENCE            1.107      0.342      3.240      0.001

 SCIENCE    ON
    FL2                0.100      0.042      2.357      0.018

 Intercepts
    SCIENCE            0.557      0.053     10.496      0.000

 Thresholds
    ENROLL$1          -1.323      0.308     -4.289      0.000

 Residual Variances
    SCIENCE            0.243      0.005     50.984      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 ENROLL     ON
    FL2                1.422
    SCIENCE            3.026
```

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ENROLL ON				
FL2	0.133	0.079	1.680	0.093
SCIENCE	0.287	0.082	3.516	0.000
SCIENCE ON				
FL2	0.146	0.062	2.350	0.019
Intercepts				
SCIENCE	1.118	0.103	10.859	0.000
Thresholds				
ENROLL$1	-0.688	0.162	-4.240	0.000
Residual Variances				
SCIENCE	0.979	0.018	54.227	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ENROLL	0.111	0.050	2.204	0.027
SCIENCE	0.021	0.018	1.175	0.240

<u>Disinterested Success → Planned STEM Major → College Enrollment.</u>

MODEL FIT INFORMATION

Number of Free Parameters	6
Loglikelihood	
H0 Value	-338.339
H0 Scaling Correction Factor for MLR	0.8432
Information Criteria	
Akaike (AIC)	688.679
Bayesian (BIC)	710.530
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	691.504

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ENROLL ON				
FL2	0.463	0.224	2.073	0.038
STEMMAJ	0.880	0.326	2.704	0.007
STEMMAJ ON				
FL2	0.024	0.040	0.583	0.560
Intercepts				
STEMMAJ	0.470	0.049	9.529	0.000
Thresholds				
ENROLL$1	-1.518	0.321	-4.724	0.000
Residual Variances				
STEMMAJ	0.247	0.003	74.866	0.000

LOGISTIC REGRESSION ODDS RATIO RESULTS

ENROLL ON	
FL2	1.590
STEMMAJ	2.411

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ENROLL ON				
FL2	0.178	0.084	2.119	0.034
STEMMAJ	0.230	0.081	2.834	0.005
STEMMAJ ON				
FL2	0.035	0.059	0.583	0.560
Intercepts				
STEMMAJ	0.944	0.096	9.884	0.000
Thresholds				
ENROLL$1	-0.799	0.167	-4.784	0.000
Residual Variances				
STEMMAJ	0.999	0.004	244.324	0.000

```
R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    ENROLL             0.088      0.043      2.013      0.044
    STEMMAJ            0.001      0.004      0.292      0.771
```

<u>Disinterested Success → Planned STEM Career → College Enrollment.</u>

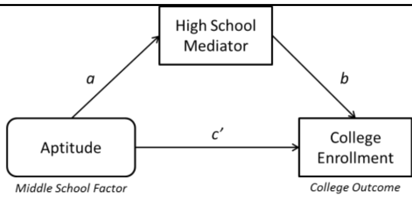
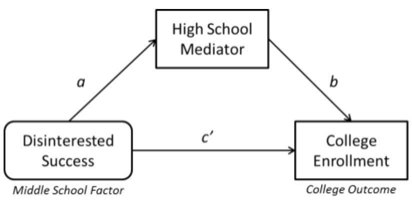
```
MODEL FIT INFORMATION

Number of Free Parameters                        6

Loglikelihood

          H0 Value                        -294.420
          H0 Scaling Correction Factor      0.9755
            for MLR

Information Criteria

          Akaike (AIC)                     600.839
          Bayesian (BIC)                   622.691
          Sample-Size Adjusted BIC         603.665
            (n* = (n + 2) / 24)


MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL2                0.474      0.230      2.059      0.039
    STEMCAR            0.545      0.393      1.385      0.166

 STEMCAR    ON
    FL2                0.021      0.034      0.631      0.528

 Intercepts
    STEMCAR            0.251      0.042      5.979      0.000

 Thresholds
    ENROLL$1          -1.756      0.314     -5.584      0.000

 Residual Variances
    STEMCAR            0.177      0.014     13.119      0.000
```

```
LOGISTIC REGRESSION ODDS RATIO RESULTS

 ENROLL     ON
    FL2                1.607
    STEMCAR            1.724


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ENROLL     ON
    FL2                0.185      0.088      2.116      0.034
    STEMCAR            0.123      0.088      1.399      0.162

 STEMCAR    ON
    FL2                0.037      0.058      0.631      0.528

 Intercepts
    STEMCAR            0.596      0.087      6.847      0.000

 Thresholds
    ENROLL$1          -0.943      0.160     -5.894      0.000

 Residual Variances
    STEMCAR            0.999      0.004    234.628      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    ENROLL             0.051      0.036      1.423      0.155
    STEMCAR            0.001      0.004      0.316      0.752
```

2. High school mediator treated as categorical variable

Except for the ***a*** coefficient, the model estimates for ***b*** and ***c'*** coefficients are the same as in the models if the high school mediators were treated as continuous (Table 27).

Model Estimates of College Enrollment from Middle School Factors with Mediation of High School Variables between (Existing mediations in bold)

Mediational Model	*High School Mediator*	*Middle School Factor → High School Mediator (a)*	*High School Mediator → College (b)*	*Middle School Factor → College Outcome (c')*	*Indirect Effect (Sobel test statistic)*
	*AP Math*	1.576**	0.322	0.595*	0.849
	AP Science	**1.209****	**0.714**$^{+}$	**0.529***	**1.854**$^{+}$
	Planned STEM Major	0.317*	0.765*	0.649**	1.695
	Planned STEM Career	0.238	0.488	0.679**	0.959
	***AP Math***	**0.679****	**0.850***	**0.342**	**2.148***
	AP Science	**0.411***	**1.107***	**0.352**	**1.937**$^{+}$
	Planned STEM Major	0.095	0.880*	0.463*	0.566
	Planned STEM Career	0.118	0.545	0.474*	0.561

** $p < 0.05$; ** $p < 0.001$; + p is marginally significant*
*Note: **a, b**, and **c'** coefficients are binary logistic regression coefficients*

APPENDIX F

MPlus Outputs of Simple Ordinal Logistic Regression Model of Selectivity of College Attended in Study 3

Model Output with Independent Variable: *Aptitude*

```
MODEL FIT INFORMATION

Number of Free Parameters                         10

Loglikelihood

          H0 Value                          -500.305
          H0 Scaling Correction Factor        1.0049
            for MLR

Information Criteria

          Akaike (AIC)                      1020.611
          Bayesian (BIC)                    1057.030
          Sample-Size Adjusted BIC          1025.320
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL1                1.279      0.134      9.536      0.000

 Thresholds
    SELORD$1          -0.641      0.177     -3.619      0.000
    SELORD$2           0.339      0.145      2.345      0.019
    SELORD$3           0.470      0.151      3.107      0.002
    SELORD$4           1.926      0.192     10.019      0.000
    SELORD$5           2.155      0.200     10.754      0.000
    SELORD$6           3.293      0.236     13.946      0.000
    SELORD$7           3.537      0.252     14.010      0.000
    SELORD$8           4.158      0.298     13.930      0.000
    SELORD$9           4.754      0.355     13.382      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL1                3.593


STANDARDIZED MODEL RESULTS
```

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL1	0.567	0.043	13.042	0.000
Thresholds				
SELORD$1	-0.291	0.082	-3.542	0.000
SELORD$2	0.154	0.063	2.427	0.015
SELORD$3	0.213	0.065	3.262	0.001
SELORD$4	0.874	0.070	12.452	0.000
SELORD$5	0.979	0.072	13.529	0.000
SELORD$6	1.496	0.081	18.379	0.000
SELORD$7	1.606	0.087	18.373	0.000
SELORD$8	1.888	0.105	17.924	0.000
SELORD$9	2.159	0.133	16.280	0.000

STDY Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL1	0.581	0.042	13.806	0.000
Thresholds				
SELORD$1	-0.291	0.082	-3.542	0.000
SELORD$2	0.154	0.063	2.427	0.015
SELORD$3	0.213	0.065	3.262	0.001
SELORD$4	0.874	0.070	12.452	0.000
SELORD$5	0.979	0.072	13.529	0.000
SELORD$6	1.496	0.081	18.379	0.000
SELORD$7	1.606	0.087	18.373	0.000
SELORD$8	1.888	0.105	17.924	0.000
SELORD$9	2.159	0.133	16.280	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.321	0.049	6.521	0.000

Model Output with Independent Variable: *Disinterested Success*

```
MODEL FIT INFORMATION

Number of Free Parameters                       10

Loglikelihood

          H0 Value                        -546.504
          H0 Scaling Correction Factor      1.0191
            for MLR

Information Criteria

          Akaike (AIC)                    1113.007
          Bayesian (BIC)                  1149.426
          Sample-Size Adjusted BIC        1117.716
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL2                0.617      0.166      3.713      0.000

 Thresholds
    SELORD$1          -1.982      0.256     -7.730      0.000
    SELORD$2          -1.183      0.236     -5.009      0.000
    SELORD$3          -1.087      0.232     -4.685      0.000
    SELORD$4           0.011      0.221      0.048      0.962
    SELORD$5           0.192      0.221      0.868      0.385
    SELORD$6           1.131      0.239      4.739      0.000
    SELORD$7           1.338      0.248      5.405      0.000
    SELORD$8           1.882      0.281      6.692      0.000
    SELORD$9           2.424      0.325      7.449      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL2                1.854


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
```

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
FL2	0.241	0.062	3.894	0.000
Thresholds				
SELORD$1	-1.061	0.124	-8.545	0.000
SELORD$2	-0.633	0.118	-5.361	0.000
SELORD$3	-0.582	0.117	-4.991	0.000
SELORD$4	0.006	0.118	0.048	0.962
SELORD$5	0.103	0.119	0.859	0.390
SELORD$6	0.605	0.135	4.490	0.000
SELORD$7	0.716	0.141	5.090	0.000
SELORD$8	1.007	0.161	6.242	0.000
SELORD$9	1.297	0.186	6.957	0.000

STDY Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL2	0.330	0.084	3.941	0.000
Thresholds				
SELORD$1	-1.061	0.124	-8.545	0.000
SELORD$2	-0.633	0.118	-5.361	0.000
SELORD$3	-0.582	0.117	-4.991	0.000
SELORD$4	0.006	0.118	0.048	0.962
SELORD$5	0.103	0.119	0.859	0.390
SELORD$6	0.605	0.135	4.490	0.000
SELORD$7	0.716	0.141	5.090	0.000
SELORD$8	1.007	0.161	6.242	0.000
SELORD$9	1.297	0.186	6.957	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.058	0.030	1.947	0.052

Model Output with Independent Variable: *Negative Emotions*

MODEL FIT INFORMATION

Number of Free Parameters	10

Loglikelihood

H0 Value	-544.751
H0 Scaling Correction Factor for MLR	0.9922

Information Criteria

Akaike (AIC)	1109.502
Bayesian (BIC)	1145.921
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	1114.211

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL3	-0.556	0.118	-4.708	0.000
Thresholds				
SELORD$1	-1.010	0.151	-6.704	0.000
SELORD$2	-0.202	0.132	-1.531	0.126
SELORD$3	-0.104	0.132	-0.788	0.431
SELORD$4	1.025	0.143	7.183	0.000
SELORD$5	1.211	0.148	8.199	0.000
SELORD$6	2.162	0.189	11.456	0.000
SELORD$7	2.368	0.200	11.867	0.000
SELORD$8	2.908	0.235	12.367	0.000
SELORD$9	3.444	0.294	11.704	0.000

LOGISTIC REGRESSION ODDS RATIO RESULTS

SELORD ON	
FL3	0.574

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				

FL3	-0.258	0.052	-4.947	0.000
Thresholds				
SELORD$1	-0.538	0.082	-6.596	0.000
SELORD$2	-0.108	0.071	-1.519	0.129
SELORD$3	-0.055	0.071	-0.784	0.433
SELORD$4	0.546	0.072	7.546	0.000
SELORD$5	0.645	0.074	8.674	0.000
SELORD$6	1.151	0.094	12.292	0.000
SELORD$7	1.261	0.099	12.701	0.000
SELORD$8	1.549	0.119	13.069	0.000
SELORD$9	1.835	0.150	12.270	0.000

STDY Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL3	-0.296	0.059	-5.042	0.000
Thresholds				
SELORD$1	-0.538	0.082	-6.596	0.000
SELORD$2	-0.108	0.071	-1.519	0.129
SELORD$3	-0.055	0.071	-0.784	0.433
SELORD$4	0.546	0.072	7.546	0.000
SELORD$5	0.645	0.074	8.674	0.000
SELORD$6	1.151	0.094	12.292	0.000
SELORD$7	1.261	0.099	12.701	0.000
SELORD$8	1.549	0.119	13.069	0.000
SELORD$9	1.835	0.150	12.270	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.066	0.027	2.473	0.013

Model Output with Independent Variable: *Engaged Concentration*

```
MODEL FIT INFORMATION

Number of Free Parameters                       10

Loglikelihood

          H0 Value                        -552.476
          H0 Scaling Correction Factor      0.9729
            for MLR

Information Criteria

          Akaike (AIC)                    1124.951
          Bayesian (BIC)                  1161.370
          Sample-Size Adjusted BIC        1129.660
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL4                0.291      0.107      2.728      0.006

 Thresholds
    SELORD$1          -1.186      0.152     -7.810      0.000
    SELORD$2          -0.406      0.131     -3.108      0.002
    SELORD$3          -0.315      0.130     -2.425      0.015
    SELORD$4           0.746      0.132      5.668      0.000
    SELORD$5           0.924      0.135      6.852      0.000
    SELORD$6           1.848      0.172     10.730      0.000
    SELORD$7           2.051      0.185     11.116      0.000
    SELORD$8           2.583      0.224     11.544      0.000
    SELORD$9           3.115      0.276     11.272      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL4                1.337


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
```

```
 SELORD     ON
    FL4                0.128      0.047      2.756      0.006

 Thresholds
    SELORD$1          -0.648      0.084     -7.707      0.000
    SELORD$2          -0.222      0.072     -3.090      0.002
    SELORD$3          -0.172      0.071     -2.414      0.016
    SELORD$4           0.408      0.071      5.730      0.000
    SELORD$5           0.505      0.073      6.940      0.000
    SELORD$6           1.010      0.093     10.918      0.000
    SELORD$7           1.121      0.099     11.308      0.000
    SELORD$8           1.413      0.121     11.701      0.000
    SELORD$9           1.703      0.150     11.360      0.000


STDY Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL4                0.159      0.057      2.774      0.006

 Thresholds
    SELORD$1          -0.648      0.084     -7.707      0.000
    SELORD$2          -0.222      0.072     -3.090      0.002
    SELORD$3          -0.172      0.071     -2.414      0.016
    SELORD$4           0.408      0.071      5.730      0.000
    SELORD$5           0.505      0.073      6.940      0.000
    SELORD$6           1.010      0.093     10.918      0.000
    SELORD$7           1.121      0.099     11.308      0.000
    SELORD$8           1.413      0.121     11.701      0.000
    SELORD$9           1.703      0.150     11.360      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.016      0.012      1.378      0.168
```

APPENDIX G

MPlus Outputs of Modeling Selectivity of College Attended from Middle School Performance-Engagement Factor with Mediation from High School Variables in Study 3

1. High school mediator treated as continuous variable (From Table 29)

Aptitude → AP Math → Selectivity of College Attended.

```
MODEL FIT INFORMATION

Number of Free Parameters                       14

Loglikelihood

          H0 Value                        -641.840
          H0 Scaling Correction Factor      0.9510
            for MLR

Information Criteria

          Akaike (AIC)                    1311.680
          Bayesian (BIC)                  1362.667
          Sample-Size Adjusted BIC        1318.273
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL1                1.023      0.148      6.929      0.000
    MATH               1.038      0.269      3.862      0.000

 MATH       ON
    FL1                0.292      0.020     14.953      0.000

 Intercepts
    MATH               0.232      0.030      7.735      0.000

 Thresholds
    SELORD$1          -0.450      0.176     -2.559      0.011
    SELORD$2           0.545      0.148      3.678      0.000
    SELORD$3           0.678      0.154      4.413      0.000
    SELORD$4           2.205      0.203     10.885      0.000
    SELORD$5           2.454      0.211     11.631      0.000
    SELORD$6           3.656      0.259     14.117      0.000
    SELORD$7           3.905      0.274     14.244      0.000
    SELORD$8           4.536      0.312     14.532      0.000
    SELORD$9           5.137      0.367     14.013      0.000
```

```
 Residual Variances
    MATH               0.169      0.011     14.839      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL1                2.780
    MATH               2.825


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL1                0.440      0.055      7.997      0.000
    MATH               0.229      0.057      4.050      0.000

 MATH       ON
    FL1                0.570      0.041     13.867      0.000

 Intercepts
    MATH               0.463      0.061      7.602      0.000

 Thresholds
    SELORD$1          -0.199      0.079     -2.521      0.012
    SELORD$2           0.240      0.062      3.900      0.000
    SELORD$3           0.299      0.063      4.743      0.000
    SELORD$4           0.972      0.068     14.258      0.000
    SELORD$5           1.082      0.069     15.568      0.000
    SELORD$6           1.612      0.081     19.814      0.000
    SELORD$7           1.722      0.087     19.900      0.000
    SELORD$8           2.000      0.101     19.778      0.000
    SELORD$9           2.264      0.127     17.805      0.000

 Residual Variances
    MATH               0.675      0.047     14.417      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.361      0.050      7.194      0.000
    MATH               0.325      0.047      6.934      0.000
```

Aptitude → AP Science → Selectivity of College Attended.

```
MODEL FIT INFORMATION

Number of Free Parameters                       14

Loglikelihood

          H0 Value                        -653.208
          H0 Scaling Correction Factor      0.9512
            for MLR

Information Criteria

          Akaike (AIC)                    1334.417
          Bayesian (BIC)                  1385.403
          Sample-Size Adjusted BIC        1341.010
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL1                1.011      0.145      6.958      0.000
    SCIENCE            1.371      0.287      4.784      0.000

 SCIENCE    ON
    FL1                0.245      0.022     11.262      0.000

 Intercepts
    SCIENCE            0.236      0.031      7.624      0.000

 Thresholds
    SELORD$1          -0.411      0.179     -2.298      0.022
    SELORD$2           0.603      0.146      4.125      0.000
    SELORD$3           0.743      0.153      4.845      0.000
    SELORD$4           2.355      0.213     11.036      0.000
    SELORD$5           2.615      0.226     11.565      0.000
    SELORD$6           3.849      0.272     14.142      0.000
    SELORD$7           4.104      0.288     14.230      0.000
    SELORD$8           4.745      0.329     14.406      0.000
    SELORD$9           5.349      0.389     13.753      0.000

 Residual Variances
    SCIENCE            0.191      0.011     17.857      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL1                2.748
    SCIENCE            3.940
```

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL1	0.426	0.054	7.875	0.000
SCIENCE	0.295	0.057	5.165	0.000
SCIENCE ON				
FL1	0.479	0.045	10.603	0.000
Intercepts				
SCIENCE	0.473	0.061	7.749	0.000
Thresholds				
SELORD$1	-0.177	0.078	-2.273	0.023
SELORD$2	0.260	0.059	4.396	0.000
SELORD$3	0.320	0.061	5.248	0.000
SELORD$4	1.016	0.069	14.714	0.000
SELORD$5	1.128	0.072	15.586	0.000
SELORD$6	1.660	0.082	20.218	0.000
SELORD$7	1.770	0.087	20.269	0.000
SELORD$8	2.046	0.102	20.045	0.000
SELORD$9	2.307	0.128	17.978	0.000
Residual Variances				
SCIENCE	0.770	0.043	17.790	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.388	0.049	7.929	0.000
SCIENCE	0.230	0.043	5.302	0.000

Aptitude → Planned STEM Major → Selectivity of College Attended.

MODEL FIT INFORMATION

Number of Free Parameters 14

Loglikelihood

H0 Value	-693.887
H0 Scaling Correction Factor for MLR	0.9328

```
Information Criteria

          Akaike (AIC)                    1415.774
          Bayesian (BIC)                  1466.761
          Sample-Size Adjusted BIC        1422.367
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL1                1.250      0.139      9.022      0.000
    STEMMAJ            0.767      0.209      3.679      0.000

 STEMMAJ    ON
    FL1                0.077      0.030      2.586      0.010

 Intercepts
    STEMMAJ            0.376      0.040      9.492      0.000

 Thresholds
    SELORD$1          -0.357      0.200     -1.788      0.074
    SELORD$2           0.634      0.174      3.641      0.000
    SELORD$3           0.766      0.180      4.244      0.000
    SELORD$4           2.258      0.211     10.691      0.000
    SELORD$5           2.493      0.218     11.454      0.000
    SELORD$6           3.663      0.248     14.777      0.000
    SELORD$7           3.915      0.265     14.793      0.000
    SELORD$8           4.554      0.313     14.559      0.000
    SELORD$9           5.165      0.357     14.471      0.000

 Residual Variances
    STEMMAJ            0.242      0.005     45.534      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL1                3.491
    STEMMAJ            2.154


STANDARDIZED MODEL RESULTS

STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL1                0.542      0.046     11.879      0.000
    STEMMAJ            0.169      0.045      3.743      0.000
```

```
 STEMMAJ    ON
    FL1                0.151      0.059      2.572      0.010

 Intercepts
    STEMMAJ            0.757      0.076      9.904      0.000

 Thresholds
    SELORD$1          -0.159      0.090     -1.762      0.078
    SELORD$2           0.282      0.073      3.855      0.000
    SELORD$3           0.340      0.075      4.556      0.000
    SELORD$4           1.003      0.074     13.549      0.000
    SELORD$5           1.108      0.075     14.712      0.000
    SELORD$6           1.628      0.082     19.932      0.000
    SELORD$7           1.740      0.087     19.888      0.000
    SELORD$8           2.023      0.106     19.120      0.000
    SELORD$9           2.295      0.129     17.842      0.000

 Residual Variances
    STEMMAJ            0.977      0.018     55.150      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.350      0.049      7.160      0.000
    STEMMAJ            0.023      0.018      1.286      0.198
```

Aptitude → Planned STEM Career → Selectivity of College Attended.

```
MODEL FIT INFORMATION

Number of Free Parameters                       14

Loglikelihood

          H0 Value                        -653.935
          H0 Scaling Correction Factor      0.9949
            for MLR

Information Criteria

          Akaike (AIC)                    1335.870
          Bayesian (BIC)                  1386.856
          Sample-Size Adjusted BIC        1342.463
            (n* = (n + 2) / 24)



MODEL RESULTS
```

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL1                1.265      0.135      9.350      0.000
    STEMCAR            0.412      0.242      1.702      0.089

 STEMCAR    ON
    FL1                0.042      0.027      1.525      0.127

 Intercepts
    STEMCAR            0.192      0.034      5.731      0.000

 Thresholds
    SELORD$1          -0.562      0.185     -3.039      0.002
    SELORD$2           0.419      0.154      2.723      0.006
    SELORD$3           0.550      0.161      3.418      0.001
    SELORD$4           2.009      0.196     10.263      0.000
    SELORD$5           2.238      0.203     11.023      0.000
    SELORD$6           3.384      0.234     14.484      0.000
    SELORD$7           3.632      0.249     14.594      0.000
    SELORD$8           4.261      0.298     14.286      0.000
    SELORD$9           4.861      0.352     13.811      0.000

 Residual Variances
    STEMCAR            0.176      0.013     13.069      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL1                3.543
    STEMCAR            1.510


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL1                0.559      0.044     12.562      0.000
    STEMCAR            0.078      0.046      1.698      0.089

 STEMCAR    ON
    FL1                0.096      0.063      1.529      0.126

 Intercepts
    STEMCAR            0.457      0.069      6.574      0.000

 Thresholds
    SELORD$1          -0.254      0.085     -2.987      0.003
    SELORD$2           0.190      0.067      2.826      0.005
    SELORD$3           0.249      0.069      3.595      0.000
```

SELORD$4	0.909	0.072	12.695	0.000
SELORD$5	1.013	0.073	13.809	0.000
SELORD$6	1.531	0.080	19.092	0.000
SELORD$7	1.643	0.086	19.195	0.000
SELORD$8	1.928	0.105	18.370	0.000
SELORD$9	2.199	0.131	16.817	0.000
Residual Variances				
STEMCAR	0.991	0.012	81.833	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.327	0.049	6.684	0.000
STEMCAR	0.009	0.012	0.764	0.445

<u>Disinterested Success → AP Math → Selectivity of College Attended.</u>

MODEL FIT INFORMATION

Number of Free Parameters 14

Loglikelihood

H0 Value	-712.271
H0 Scaling Correction Factor for MLR	0.9483

Information Criteria

Akaike (AIC)	1452.541
Bayesian (BIC)	1503.528
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	1459.134

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL2	0.404	0.154	2.631	0.009
MATH	1.840	0.252	7.317	0.000
MATH ON				
FL2	0.158	0.041	3.812	0.000

```
 Intercepts
    MATH               0.652      0.051     12.738      0.000

 Thresholds
    SELORD$1          -1.148      0.251     -4.574      0.000
    SELORD$2          -0.262      0.241     -1.088      0.276
    SELORD$3          -0.151      0.239     -0.632      0.527
    SELORD$4           1.177      0.253      4.655      0.000
    SELORD$5           1.403      0.250      5.617      0.000
    SELORD$6           2.508      0.286      8.766      0.000
    SELORD$7           2.736      0.294      9.301      0.000
    SELORD$8           3.321      0.319     10.395      0.000
    SELORD$9           3.890      0.363     10.721      0.000

 Residual Variances
    MATH               0.237      0.007     35.655      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL2                1.498
    MATH               6.299


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL2                0.141      0.053      2.677      0.007
    MATH               0.441      0.049      9.028      0.000

 MATH       ON
    FL2                0.230      0.061      3.780      0.000

 Intercepts
    MATH               1.304      0.101     12.866      0.000

 Thresholds
    SELORD$1          -0.551      0.119     -4.641      0.000
    SELORD$2          -0.126      0.115     -1.089      0.276
    SELORD$3          -0.072      0.114     -0.632      0.527
    SELORD$4           0.564      0.120      4.711      0.000
    SELORD$5           0.673      0.118      5.713      0.000
    SELORD$6           1.203      0.132      9.117      0.000
    SELORD$7           1.312      0.136      9.669      0.000
    SELORD$8           1.593      0.149     10.669      0.000
    SELORD$9           1.866      0.170     10.959      0.000

 Residual Variances
    MATH               0.947      0.028     33.941      0.000
```

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.243	0.046	5.325	0.000
MATH	0.053	0.028	1.890	0.059

Disinterested Success → AP Science → Selectivity of College Attended.

MODEL FIT INFORMATION

Number of Free Parameters 14

Loglikelihood

H0 Value	-709.569
H0 Scaling Correction Factor for MLR	0.9354

Information Criteria

Akaike (AIC)	1447.139
Bayesian (BIC)	1498.126
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	1453.732

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL2	0.490	0.132	3.702	0.000
SCIENCE	2.047	0.268	7.635	0.000
SCIENCE ON				
FL2	0.100	0.042	2.357	0.018
Intercepts				
SCIENCE	0.557	0.053	10.496	0.000
Thresholds				
SELORD$1	-1.264	0.231	-5.466	0.000
SELORD$2	-0.356	0.219	-1.621	0.105
SELORD$3	-0.237	0.218	-1.086	0.277
SELORD$4	1.183	0.240	4.928	0.000
SELORD$5	1.418	0.248	5.712	0.000
SELORD$6	2.545	0.286	8.896	0.000
SELORD$7	2.777	0.297	9.362	0.000
SELORD$8	3.371	0.327	10.305	0.000

```
    SELORD$9           3.942      0.379     10.393      0.000

 Residual Variances
    SCIENCE            0.243      0.005     50.984      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL2                1.632
    SCIENCE            7.743


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL2                0.167      0.044      3.763      0.000
    SCIENCE            0.478      0.048      9.878      0.000

 SCIENCE    ON
    FL2                0.146      0.062      2.350      0.019

 Intercepts
    SCIENCE            1.118      0.103     10.859      0.000

 Thresholds
    SELORD$1          -0.592      0.107     -5.511      0.000
    SELORD$2          -0.166      0.103     -1.614      0.107
    SELORD$3          -0.111      0.103     -1.083      0.279
    SELORD$4           0.554      0.107      5.151      0.000
    SELORD$5           0.664      0.110      6.048      0.000
    SELORD$6           1.191      0.122      9.760      0.000
    SELORD$7           1.300      0.126     10.283      0.000
    SELORD$8           1.578      0.141     11.209      0.000
    SELORD$9           1.845      0.163     11.296      0.000

 Residual Variances
    SCIENCE            0.979      0.018     54.227      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.279      0.047      5.883      0.000
    SCIENCE            0.021      0.018      1.175      0.240
```

Disinterested Success → Planned STEM Major → Selectivity of College Attended.

```
MODEL FIT INFORMATION

Number of Free Parameters                       14

Loglikelihood

          H0 Value                        -739.427
          H0 Scaling Correction Factor      0.9385
            for MLR

Information Criteria

          Akaike (AIC)                    1506.853
          Bayesian (BIC)                  1557.840
          Sample-Size Adjusted BIC        1513.446
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL2                0.600      0.161      3.728      0.000
    STEMMAJ            0.965      0.218      4.430      0.000

 STEMMAJ    ON
    FL2                0.024      0.040      0.583      0.560

 Intercepts
    STEMMAJ            0.470      0.049      9.529      0.000

 Thresholds
    SELORD$1          -1.596      0.267     -5.966      0.000
    SELORD$2          -0.773      0.251     -3.076      0.002
    SELORD$3          -0.674      0.248     -2.715      0.007
    SELORD$4           0.470      0.245      1.917      0.055
    SELORD$5           0.659      0.246      2.683      0.007
    SELORD$6           1.638      0.260      6.288      0.000
    SELORD$7           1.853      0.267      6.947      0.000
    SELORD$8           2.414      0.299      8.084      0.000
    SELORD$9           2.972      0.327      9.079      0.000

 Residual Variances
    STEMMAJ            0.247      0.003     74.866      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL2                1.822
    STEMMAJ            2.624
```

```
STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL2                0.226      0.059      3.867      0.000
    STEMMAJ            0.249      0.053      4.680      0.000

 STEMMAJ    ON
    FL2                0.035      0.059      0.583      0.560

 Intercepts
    STEMMAJ            0.944      0.096      9.884      0.000

 Thresholds
    SELORD$1          -0.827      0.134     -6.150      0.000
    SELORD$2          -0.400      0.128     -3.128      0.002
    SELORD$3          -0.349      0.127     -2.755      0.006
    SELORD$4           0.244      0.128      1.898      0.058
    SELORD$5           0.342      0.129      2.646      0.008
    SELORD$6           0.849      0.140      6.077      0.000
    SELORD$7           0.960      0.144      6.682      0.000
    SELORD$8           1.251      0.162      7.728      0.000
    SELORD$9           1.540      0.180      8.571      0.000

 Residual Variances
    STEMMAJ            0.999      0.004    244.324      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.117      0.035      3.326      0.001
    STEMMAJ            0.001      0.004      0.292      0.771
```

<u>Disinterested Success → Planned STEM Career → Selectivity of College Attended.</u>

```
MODEL FIT INFORMATION

Number of Free Parameters                         14

Loglikelihood

          H0 Value                          -699.671
          H0 Scaling Correction Factor        0.9977
            for MLR
```

Information Criteria

Akaike (AIC)	1427.343
Bayesian (BIC)	1478.330
Sample-Size Adjusted BIC	1433.936
(n* = (n + 2) / 24)	

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL2	0.606	0.162	3.743	0.000
STEMCAR	0.613	0.262	2.338	0.019
STEMCAR ON				
FL2	0.021	0.034	0.631	0.528
Intercepts				
STEMCAR	0.251	0.042	5.979	0.000
Thresholds				
SELORD$1	-1.845	0.260	-7.096	0.000
SELORD$2	-1.039	0.241	-4.305	0.000
SELORD$3	-0.943	0.238	-3.965	0.000
SELORD$4	0.163	0.229	0.713	0.476
SELORD$5	0.345	0.228	1.511	0.131
SELORD$6	1.296	0.241	5.378	0.000
SELORD$7	1.507	0.247	6.111	0.000
SELORD$8	2.062	0.285	7.235	0.000
SELORD$9	2.611	0.325	8.033	0.000
Residual Variances				
STEMCAR	0.177	0.014	13.119	0.000

LOGISTIC REGRESSION ODDS RATIO RESULTS

SELORD ON	
FL2	1.833
STEMCAR	1.846

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL2	0.234	0.060	3.910	0.000
STEMCAR	0.137	0.058	2.370	0.018

```
 STEMCAR    ON
    FL2                0.037      0.058      0.631      0.528

 Intercepts
    STEMCAR            0.596      0.087      6.847      0.000

 Thresholds
    SELORD$1          -0.978      0.129     -7.587      0.000
    SELORD$2          -0.551      0.123     -4.494      0.000
    SELORD$3          -0.500      0.121     -4.125      0.000
    SELORD$4           0.087      0.122      0.708      0.479
    SELORD$5           0.183      0.123      1.489      0.136
    SELORD$6           0.687      0.134      5.113      0.000
    SELORD$7           0.799      0.138      5.769      0.000
    SELORD$8           1.093      0.160      6.809      0.000
    SELORD$9           1.384      0.183      7.550      0.000

 Residual Variances
    STEMCAR            0.999      0.004    234.628      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.076      0.032      2.367      0.018
    STEMCAR            0.001      0.004      0.316      0.752
```

<u>Negative Emotions → AP Math → Selectivity of College Attended.</u>

```
MODEL FIT INFORMATION

Number of Free Parameters                       14

Loglikelihood

          H0 Value                        -710.980
          H0 Scaling Correction Factor      0.9297
            for MLR

Information Criteria

          Akaike (AIC)                    1449.960
          Bayesian (BIC)                  1500.946
          Sample-Size Adjusted BIC        1456.553
            (n* = (n + 2) / 24)
```

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL3               -0.448      0.125     -3.579      0.000
    MATH               1.846      0.244      7.551      0.000

 MATH       ON
    FL3               -0.120      0.032     -3.703      0.000

 Intercepts
    MATH               0.413      0.036     11.405      0.000

 Thresholds
    SELORD$1          -0.474      0.163     -2.910      0.004
    SELORD$2           0.428      0.150      2.842      0.004
    SELORD$3           0.543      0.153      3.557      0.000
    SELORD$4           1.910      0.193      9.923      0.000
    SELORD$5           2.145      0.198     10.836      0.000
    SELORD$6           3.270      0.254     12.883      0.000
    SELORD$7           3.498      0.263     13.277      0.000
    SELORD$8           4.082      0.288     14.155      0.000
    SELORD$9           4.646      0.342     13.590      0.000

 Residual Variances
    MATH               0.239      0.006     40.722      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL3                0.639
    MATH               6.334


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL3               -0.185      0.050     -3.719      0.000
    MATH               0.438      0.046      9.488      0.000

 MATH       ON
    FL3               -0.208      0.057     -3.664      0.000

 Intercepts
    MATH               0.826      0.072     11.391      0.000

 Thresholds
    SELORD$1          -0.225      0.079     -2.850      0.004
```

```
    SELORD$2           0.203      0.068      2.971      0.003
    SELORD$3           0.258      0.069      3.756      0.000
    SELORD$4           0.906      0.073     12.365      0.000
    SELORD$5           1.018      0.073     13.888      0.000
    SELORD$6           1.551      0.091     17.038      0.000
    SELORD$7           1.659      0.095     17.458      0.000
    SELORD$8           1.937      0.108     17.920      0.000
    SELORD$9           2.204      0.135     16.385      0.000

 Residual Variances
    MATH               0.957      0.024     40.411      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.260      0.047      5.503      0.000
    MATH               0.043      0.024      1.832      0.067
```

<u>Negative Emotions → AP Science → Selectivity of College Attended.</u>

```
MODEL FIT INFORMATION

Number of Free Parameters                       14

Loglikelihood

          H0 Value                        -699.149
          H0 Scaling Correction Factor      0.9315
            for MLR

Information Criteria

          Akaike (AIC)                    1426.299
          Bayesian (BIC)                  1477.286
          Sample-Size Adjusted BIC        1432.892
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL3               -0.226      0.130     -1.734      0.083
    SCIENCE            1.966      0.283      6.935      0.000

 SCIENCE    ON
    FL3               -0.197      0.028     -7.130      0.000
```

```
 Intercepts
    SCIENCE            0.317      0.030     10.676      0.000

 Thresholds
    SELORD$1          -0.631      0.157     -4.021      0.000
    SELORD$2           0.261      0.142      1.838      0.066
    SELORD$3           0.377      0.144      2.620      0.009
    SELORD$4           1.777      0.183      9.733      0.000
    SELORD$5           2.009      0.194     10.373      0.000
    SELORD$6           3.118      0.241     12.921      0.000
    SELORD$7           3.347      0.252     13.264      0.000
    SELORD$8           3.929      0.284     13.839      0.000
    SELORD$9           4.488      0.344     13.035      0.000

 Residual Variances
    SCIENCE            0.219      0.008     26.561      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL3                0.798
    SCIENCE            7.141


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL3               -0.094      0.054     -1.730      0.084
    SCIENCE            0.466      0.054      8.584      0.000

 SCIENCE    ON
    FL3               -0.344      0.049     -6.975      0.000

 Intercepts
    SCIENCE            0.637      0.058     10.947      0.000

 Thresholds
    SELORD$1          -0.300      0.076     -3.947      0.000
    SELORD$2           0.124      0.066      1.880      0.060
    SELORD$3           0.179      0.066      2.707      0.007
    SELORD$4           0.845      0.071     11.864      0.000
    SELORD$5           0.955      0.074     12.886      0.000
    SELORD$6           1.483      0.089     16.672      0.000
    SELORD$7           1.592      0.093     17.043      0.000
    SELORD$8           1.868      0.109     17.210      0.000
    SELORD$9           2.135      0.137     15.629      0.000

 Residual Variances
    SCIENCE            0.882      0.034     25.984      0.000
```

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.256	0.046	5.522	0.000
SCIENCE	0.118	0.034	3.488	0.000

Negative Emotions → Planned STEM Major → Selectivity of College Attended.

MODEL FIT INFORMATION

Number of Free Parameters 14

Loglikelihood

H0 Value	-736.764
H0 Scaling Correction Factor for MLR	0.9226

Information Criteria

Akaike (AIC)	1501.527
Bayesian (BIC)	1552.514
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	1508.120

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL3	-0.575	0.122	-4.708	0.000
STEMMAJ	1.006	0.216	4.664	0.000
STEMMAJ ON				
FL3	-0.020	0.033	-0.612	0.541
Intercepts				
STEMMAJ	0.432	0.038	11.359	0.000
Thresholds				
SELORD$1	-0.627	0.173	-3.615	0.000
SELORD$2	0.212	0.158	1.339	0.180
SELORD$3	0.314	0.160	1.964	0.050
SELORD$4	1.498	0.178	8.414	0.000
SELORD$5	1.695	0.185	9.159	0.000
SELORD$6	2.694	0.220	12.229	0.000
SELORD$7	2.909	0.228	12.742	0.000
SELORD$8	3.467	0.262	13.243	0.000

```
    SELORD$9           4.018      0.303     13.258      0.000

 Residual Variances
    STEMMAJ            0.247      0.003     74.677      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL3                0.563
    STEMMAJ            2.734


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL3               -0.257      0.052     -4.974      0.000
    STEMMAJ            0.256      0.051      5.030      0.000

 STEMMAJ    ON
    FL3               -0.036      0.058     -0.612      0.541

 Intercepts
    STEMMAJ            0.869      0.072     12.016      0.000

 Thresholds
    SELORD$1          -0.321      0.092     -3.495      0.000
    SELORD$2           0.109      0.080      1.362      0.173
    SELORD$3           0.161      0.080      2.015      0.044
    SELORD$4           0.768      0.081      9.534      0.000
    SELORD$5           0.869      0.083     10.496      0.000
    SELORD$6           1.380      0.096     14.355      0.000
    SELORD$7           1.491      0.100     14.901      0.000
    SELORD$8           1.777      0.117     15.124      0.000
    SELORD$9           2.059      0.141     14.594      0.000

 Residual Variances
    STEMMAJ            0.999      0.004    239.247      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.136      0.038      3.584      0.000
    STEMMAJ            0.001      0.004      0.306      0.760
```

Negative Emotions → Planned STEM Career → Selectivity of College Attended.

```
MODEL FIT INFORMATION

Number of Free Parameters                       14

Loglikelihood

          H0 Value                        -697.586
          H0 Scaling Correction Factor      0.9867
            for MLR

Information Criteria

          Akaike (AIC)                    1423.173
          Bayesian (BIC)                  1474.159
          Sample-Size Adjusted BIC        1429.765
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL3               -0.566      0.120     -4.712      0.000
    STEMCAR            0.662      0.256      2.584      0.010

 STEMCAR    ON
    FL3                0.004      0.029      0.139      0.889

 Intercepts
    STEMCAR            0.233      0.033      7.059      0.000

 Thresholds
    SELORD$1          -0.871      0.161     -5.400      0.000
    SELORD$2          -0.055      0.143     -0.382      0.703
    SELORD$3           0.045      0.144      0.309      0.757
    SELORD$4           1.185      0.156      7.592      0.000
    SELORD$5           1.373      0.161      8.534      0.000
    SELORD$6           2.338      0.194     12.029      0.000
    SELORD$7           2.549      0.202     12.630      0.000
    SELORD$8           3.100      0.243     12.779      0.000
    SELORD$9           3.643      0.297     12.271      0.000

 Residual Variances
    STEMCAR            0.177      0.014     13.121      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL3                0.568
    STEMCAR            1.939
```

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL3	-0.260	0.052	-4.949	0.000
STEMCAR	0.147	0.056	2.630	0.009
STEMCAR ON				
FL3	0.008	0.060	0.139	0.889
Intercepts				
STEMCAR	0.554	0.064	8.684	0.000
Thresholds				
SELORD$1	-0.459	0.087	-5.251	0.000
SELORD$2	-0.029	0.076	-0.381	0.703
SELORD$3	0.024	0.076	0.310	0.757
SELORD$4	0.624	0.077	8.157	0.000
SELORD$5	0.723	0.078	9.248	0.000
SELORD$6	1.231	0.093	13.236	0.000
SELORD$7	1.342	0.097	13.845	0.000
SELORD$8	1.632	0.118	13.830	0.000
SELORD$9	1.918	0.147	13.065	0.000
Residual Variances				
STEMCAR	1.000	0.001	983.808	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.088	0.031	2.856	0.004
STEMCAR	0.000	0.001	0.070	0.944

<u>Engaged Concentration → AP Math → Selectivity of College Attended.</u>

MODEL FIT INFORMATION

Number of Free Parameters 14

Loglikelihood

H0 Value	-720.662
H0 Scaling Correction Factor for MLR	0.9126

```
Information Criteria

          Akaike (AIC)                    1469.323
          Bayesian (BIC)                  1520.310
          Sample-Size Adjusted BIC        1475.916
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL4                0.242      0.101      2.385      0.017
    MATH               1.925      0.248      7.753      0.000

 MATH       ON
    FL4                0.054      0.038      1.431      0.152

 Intercepts
    MATH               0.474      0.035     13.627      0.000

 Thresholds
    SELORD$1          -0.578      0.161     -3.597      0.000
    SELORD$2           0.301      0.148      2.033      0.042
    SELORD$3           0.411      0.150      2.747      0.006
    SELORD$4           1.728      0.172     10.065      0.000
    SELORD$5           1.955      0.173     11.281      0.000
    SELORD$6           3.057      0.226     13.527      0.000
    SELORD$7           3.283      0.236     13.902      0.000
    SELORD$8           3.860      0.263     14.650      0.000
    SELORD$9           4.421      0.313     14.118      0.000

 Residual Variances
    MATH               0.248      0.003     95.441      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL4                1.274
    MATH               6.857


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL4                0.094      0.039      2.395      0.017
    MATH               0.465      0.047      9.920      0.000
```

```
 MATH       ON
    FL4                0.087      0.061      1.429      0.153

 Intercepts
    MATH               0.947      0.070     13.619      0.000

 Thresholds
    SELORD$1          -0.279      0.079     -3.518      0.000
    SELORD$2           0.146      0.070      2.085      0.037
    SELORD$3           0.199      0.070      2.838      0.005
    SELORD$4           0.834      0.069     12.029      0.000
    SELORD$5           0.944      0.068     13.889      0.000
    SELORD$6           1.477      0.086     17.111      0.000
    SELORD$7           1.585      0.091     17.442      0.000
    SELORD$8           1.864      0.106     17.597      0.000
    SELORD$9           2.135      0.132     16.185      0.000

 Residual Variances
    MATH               0.992      0.011     93.927      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.233      0.044      5.246      0.000
    MATH               0.008      0.011      0.714      0.475
```

<u>Engaged Concentration → AP Science → Selectivity of College Attended.</u>

```
MODEL FIT INFORMATION

Number of Free Parameters                       14

Loglikelihood

          H0 Value                        -714.059
          H0 Scaling Correction Factor      0.9111
            for MLR

Information Criteria

          Akaike (AIC)                    1456.117
          Bayesian (BIC)                  1507.104
          Sample-Size Adjusted BIC        1462.710
            (n* = (n + 2) / 24)
```

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL4	0.171	0.105	1.633	0.102
SCIENCE	2.071	0.265	7.826	0.000
SCIENCE ON				
FL4	0.096	0.035	2.759	0.006
Intercepts				
SCIENCE	0.414	0.032	12.879	0.000
Thresholds				
SELORD$1	-0.647	0.161	-4.008	0.000
SELORD$2	0.244	0.150	1.626	0.104
SELORD$3	0.358	0.153	2.348	0.019
SELORD$4	1.746	0.194	9.021	0.000
SELORD$5	1.979	0.206	9.621	0.000
SELORD$6	3.089	0.250	12.352	0.000
SELORD$7	3.318	0.261	12.690	0.000
SELORD$8	3.898	0.292	13.348	0.000
SELORD$9	4.457	0.347	12.843	0.000
Residual Variances				
SCIENCE	0.242	0.005	50.426	0.000

LOGISTIC REGRESSION ODDS RATIO RESULTS

SELORD ON	
FL4	1.186
SCIENCE	7.929

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL4	0.066	0.040	1.640	0.101
SCIENCE	0.491	0.047	10.363	0.000
SCIENCE ON				
FL4	0.155	0.057	2.745	0.006
Intercepts				
SCIENCE	0.830	0.061	13.588	0.000
Thresholds				
SELORD$1	-0.308	0.079	-3.902	0.000

SELORD$2	0.116	0.070	1.664	0.096
SELORD$3	0.170	0.070	2.429	0.015
SELORD$4	0.831	0.076	11.001	0.000
SELORD$5	0.941	0.079	11.952	0.000
SELORD$6	1.470	0.091	16.120	0.000
SELORD$7	1.578	0.096	16.495	0.000
SELORD$8	1.855	0.110	16.826	0.000
SELORD$9	2.120	0.136	15.584	0.000
Residual Variances				
SCIENCE	0.976	0.018	55.604	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.255	0.047	5.399	0.000
SCIENCE	0.024	0.018	1.372	0.170

<u>Engaged Concentration → Planned STEM Major → Selectivity of College Attended.</u>

MODEL FIT INFORMATION

Number of Free Parameters 14

Loglikelihood

H0 Value	-744.741
H0 Scaling Correction Factor for MLR	0.9078

Information Criteria

Akaike (AIC)	1517.482
Bayesian (BIC)	1568.468
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	1524.075

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL4	0.313	0.108	2.909	0.004
STEMMAJ	1.002	0.217	4.614	0.000
STEMMAJ ON				
FL4	-0.008	0.036	-0.229	0.819

```
 Intercepts
    STEMMAJ            0.451      0.035     13.045      0.000

 Thresholds
    SELORD$1          -0.798      0.174     -4.574      0.000
    SELORD$2           0.010      0.161      0.064      0.949
    SELORD$3           0.105      0.162      0.652      0.515
    SELORD$4           1.219      0.168      7.239      0.000
    SELORD$5           1.406      0.171      8.201      0.000
    SELORD$6           2.373      0.201     11.825      0.000
    SELORD$7           2.585      0.210     12.329      0.000
    SELORD$8           3.134      0.246     12.753      0.000
    SELORD$9           3.679      0.279     13.179      0.000

 Residual Variances
    STEMMAJ            0.247      0.003     77.916      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL4                1.368
    STEMMAJ            2.723


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL4                0.133      0.045      2.939      0.003
    STEMMAJ            0.263      0.053      4.957      0.000

 STEMMAJ    ON
    FL4               -0.013      0.058     -0.229      0.819

 Intercepts
    STEMMAJ            0.907      0.065     14.033      0.000

 Thresholds
    SELORD$1          -0.421      0.096     -4.400      0.000
    SELORD$2           0.005      0.085      0.064      0.949
    SELORD$3           0.056      0.085      0.656      0.512
    SELORD$4           0.643      0.083      7.785      0.000
    SELORD$5           0.741      0.083      8.898      0.000
    SELORD$6           1.251      0.096     13.036      0.000
    SELORD$7           1.363      0.101     13.545      0.000
    SELORD$8           1.652      0.120     13.795      0.000
    SELORD$9           1.939      0.141     13.748      0.000

 Residual Variances
    STEMMAJ            1.000      0.002    640.555      0.000
```

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD	0.086	0.030	2.869	0.004
STEMMAJ	0.000	0.002	0.115	0.909

<u>Engaged Concentration → Planned STEM Career → Selectivity of College Attended.</u>

MODEL FIT INFORMATION

```
Number of Free Parameters                        14

Loglikelihood

          H0 Value                         -705.167
          H0 Scaling Correction Factor       0.9655
            for MLR

Information Criteria

          Akaike (AIC)                     1438.333
          Bayesian (BIC)                   1489.320
          Sample-Size Adjusted BIC         1444.926
            (n* = (n + 2) / 24)
```

MODEL RESULTS

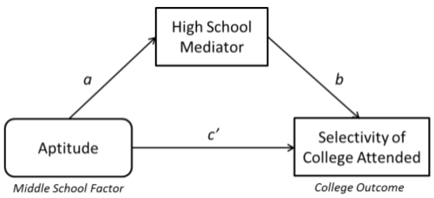
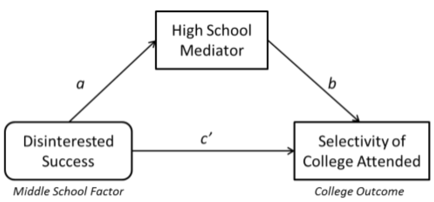
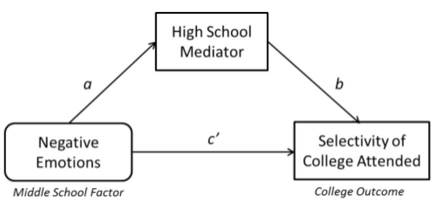
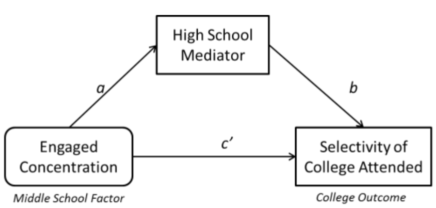
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SELORD ON				
FL4	0.302	0.111	2.727	0.006
STEMCAR	0.653	0.260	2.516	0.012
STEMCAR ON				
FL4	-0.022	0.029	-0.760	0.448
Intercepts				
STEMCAR	0.241	0.030	8.139	0.000
Thresholds				
SELORD$1	-1.048	0.163	-6.422	0.000
SELORD$2	-0.260	0.145	-1.794	0.073
SELORD$3	-0.168	0.145	-1.156	0.248
SELORD$4	0.904	0.145	6.231	0.000
SELORD$5	1.082	0.147	7.377	0.000
SELORD$6	2.017	0.175	11.512	0.000
SELORD$7	2.225	0.184	12.114	0.000

```
    SELORD$8           2.767      0.228     12.139      0.000
    SELORD$9           3.305      0.276     11.990      0.000

 Residual Variances
    STEMCAR            0.177      0.013     13.122      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS

 SELORD     ON
    FL4                1.353
    STEMCAR            1.922


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 SELORD     ON
    FL4                0.132      0.048      2.749      0.006
    STEMCAR            0.149      0.058      2.558      0.011

 STEMCAR    ON
    FL4               -0.042      0.056     -0.760      0.447

 Intercepts
    STEMCAR            0.573      0.052     10.939      0.000

 Thresholds
    SELORD$1          -0.567      0.091     -6.240      0.000
    SELORD$2          -0.141      0.079     -1.778      0.075
    SELORD$3          -0.091      0.079     -1.149      0.250
    SELORD$4           0.489      0.076      6.416      0.000
    SELORD$5           0.585      0.077      7.625      0.000
    SELORD$6           1.091      0.091     11.935      0.000
    SELORD$7           1.203      0.096     12.513      0.000
    SELORD$8           1.496      0.120     12.513      0.000
    SELORD$9           1.787      0.147     12.186      0.000

 Residual Variances
    STEMCAR            0.998      0.005    213.037      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    SELORD             0.038      0.020      1.900      0.057
    STEMCAR            0.002      0.005      0.380      0.704
```

2. High school mediator treated as categorical variable

Except for the ***a*** coefficient, the model estimates for ***b*** and ***c'*** coefficients are the same as in the models if the high school mediators were treated as continuous (Table 29)

Model Estimates of Selectivity of College Attended from Middle School Factors with Mediation of High School Variables (Significant mediations in bold)

Mediational Model	*High School Mediator*	*Middle School Factor → High School Mediator (**a**)*	*High School Mediator → College (**b**)*	*Middle School Factor → College Outcome (**c'**)*	*Indirect Effect (Sobel test statistic)*
	***AP Math***	**1.576****	**1.038****	**1.023****	**3.531****
	AP Science	**1.209****	**1.371****	**1.011****	**4.228****
	Planned STEM Major	**0.317***	**0.767****	**1.250****	**2.043***
	Planned STEM Career	0.238	0.412	1.265**	1.146
	***AP Math***	**0.679****	**1.840****	**0.404***	**3.347****
	AP Science	**0.411***	**2.047****	**0.490***	**2.307***
	Planned STEM Major	0.095	0.965**	0.600**	0.571
	Planned STEM Career	0.118	0.225*	0.355**	0.576
	***AP Math***	**-0.505***	**1.846****	**-0.448****	**-3.117***
	AP Science	**-0.974****	**1.966****	**-0.226**	**-4.352****
	Planned STEM Major	-0.083	1.006**	-0.575**	-0.592
	Planned STEM Career	0.023	0.662*	-0.566**	0.143
	*AP Math*	0.218	1.925**	0.242*	1.430
	AP Science	**0.401***	**2.071****	**0.171**	**2.456***
	Planned STEM Major	-0.033	1.002**	0.313*	-0.223
	Planned STEM Career	-0.123	0.653*	0.302*	-0.682

** p < 0.05; ** p< 0.001; + p is marginally significant*

*Note: **a** coefficients are binary logistic regression coefficients, **b** and **c'** coefficients are ordered/ordinal logistic regression coefficients*

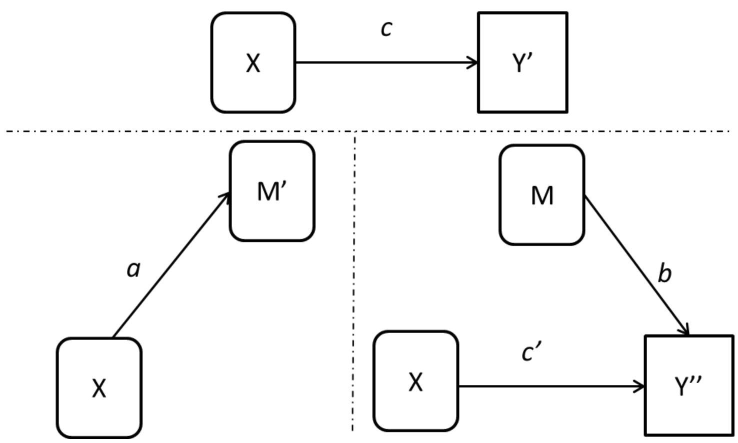
APPENDIX H

Standardizing Regression Coefficients for Mediation Significance Test

Sources:

- Liu, H., Zhang, Y., & Luo, F. (2015). Mediation Analysis for Ordinal Outcome Variables. *Quantitative Psychology Research* (pp. 429-450).
- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17, 144-158.
- Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology 1982* (pp.290-312).

MacKinnon and Dwyer (1993) presents a solution in testing for mediation with a dichotomous mediator, outcome, or both. Logistic regression creates a problem because when outcomes are dichotomous the coefficients in your mediation analyses end up being in different scales. To solve this, the overall mediation model is broken down, and represented by different equations:



$$Y' = cX + e1$$
$$M' = aX + e2$$
$$Y'' = bM + c'X + e3$$

Next is to make the coefficients comparable across the equations – multiply each coefficient by the standard deviation (SD) of the predictors in the equation, then divide by the standard deviation of the outcome variable:

$a\^{} = a * SD_X / SD_{M'}$
$b\^{} = b * SD_M / SD_{Y''}$
$c\^{} = c * SD_X / SD_{Y'}$
$c'\^{} = c' * SD_X / SD_{Y''}$

SD_X and SD_M can be derived from descriptive statistic. $SD_{M'}$, $SD_{Y'}$, $SD_{Y''}$ can be derived from the square of the variances of M', Y' and Y". MacKinnon and Dwyer (1993) provides these formulas:

$Var_{Y'} = c^2 * Var_X + p^2/3$
$Var_{M'} = a^2 * Var_X + p^2/3$
$Var_{Y''} = c'^2 * Var_X + b^2 * Var_M + 2*b*c'*Cov_{(X,M)} + \pi^2/3$

Then, the corresponding standard errors (SEs) for these comparable coefficients can be computed by:

$SE_{a^\wedge} = SE_a * SD_X / SD_{M'}$
$SE_{b^\wedge} = SE_b * SD_M / SD_{Y''}$
$SE_{c^\wedge} = SE_c * SD_X / SD_{Y'}$
$SE_{c'^\wedge} = SE_{c'} * SD_X / SD_{Y''}$

Using Sobel Test (1982) can now be conducted using the comparable coefficients and standard errors to test for significance of the amount of mediation found.