ANALYZING NON-COGNITIVE PREDICTORS OF SECONDARY AND POSTSECONDARY ACADEMIC ACHIEVEMENT

by

Brianna Cermak

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Education

Department of Educational Studies

West Lafayette, Indiana

December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Anne Traynor, Chair
Department of Educational Studies

Dr. Youli Mantzicopolous
Department of Educational Studies

Dr. Helen Patrick
Department of Educational Studies

Approved by:
Janet Alsup

TABLE OF CONTENTS

ABSTRACT

A secondary data analysis was conducted using the High School Longitudinal Study of 2009 (HSLS:09) provided by the National Center for Education Statistics to assess the relationships between academic performance indicators (high school GPA, high school mathematics achievement, and college enrollment) and perceptions of utility value, self-efficacy, effort cost, school engagement, and intelligence theories (N = 5,789). Four data collection phases occurred during HSLS:09-- 9th grade fall semester (BY), 11th grade spring semester (F1), undergraduate update in summer 2013 (U13), and a second follow-up in winter 2016 (F2).

The domain specificity and stability over time of each motivational construct was also assessed. Evaluating the domain specificity of a motivational construct helps us further understand the theoretical construction and appropriate measurement of these constructs. Motivational constructs that are more stable over time are more likely to be more effective long-term predictors of academic performance. Paired *t*-tests were conducted to evaluate the domain specificity and stability of each motivational construct. Regression models were utilized to assess motivational constructs' ability to predict academic performance.

Effort cost was the only motivational construct that was not domain specific ($t = 1.79$, $p = 0.07$). Science self-efficacy was the only motivational construct determined to be stable over time ($t = 1.19$, $p = 0.24$). School engagement, BY science efficacy, mathematics effort, and F1 science utility were significant predictors of increased academic performance for all academic performance indicators.

INTRODUCTION

Many aspects of Western society place high value on strong academic performance and degree completion. Higher levels of educational attainment are positively associated with higher income (Belfield & Bailey, 2011; Furnham & Cheng, 2013; Gitterman et al., 2015; Hout, 2012). Cutler and Lleras-Muney (2010) assessed connections between educational experience and health outcomes. They found that each additional year of education reduced the probability of smoking by 3 percentage points, of being obese by 1.4 percentage points, and of being a heavy drinker by 1.8 percentage points. A literature review by Hout (2012) revealed that college graduates are more likely to exhibit behaviors of civil engagement such as increased voter registration and increased volunteerism in local communities. As academic success is positively correlated with adaptive economic, health, and community benefits, being able to identify correlates of academic success is vital.

Cognitive skills, or skills that involve the acquisition of knowledge, manipulation of information, and reasoning (Kiely, 2014), are most commonly assessed in educational research to predict academic performance. Within the context of education, cognitive skills are often evaluated with assessments such as the ACT, GRE, intelligence tests, and other standardized testing that evaluates performance in various academic domains. The issue with solely focusing on cognitive predictors of academic performance, however, lies within the impact non-cognitive factors can have on educational success. A meta-analysis conducted by Roberts et al. (2007) found that the effect size of personality traits on educational outcomes was comparable with the effect size of IQ on similar outcomes. Kuncel and Hezlett (2010) argue that the use of standardized cognitive ability tests to predict performance can be enhanced by adding carefully selected measures of personality, values, interests, and habits to the admission system. Various studies have concluded that non-cognitive traits can contribute unique variance to the prediction of high school and college GPA (Martin, Montgomery, & Saphian, 2006; Noftle & Robins, 2007; Richardson, Abraham, & Bond, 2012; Schneider & Preckel, 2017; Wolfe & Johnson, 1995).

Research about non-cognitive constructs' ability to predict academic performance is quite abundant; however, there are some gaps in the present literature that would help make these assumptions more well-rounded. Domain specificity, or construct variability between domains, should be accounted for when testing a construct's relationship to academic performance. If a construct's domain specificity is not tested or is not appropriately considered in research, then incorrect generalizations may be created about a construct's ability to predict academic performance. A construct's stability over time is another important characteristic that should be taken into consideration when a construct is being utilized as an independent variable. Bashkov and Finney (2013) define construct stability as "the lack of change in the construct over time" (p. 289). This thesis further defines this concept as a statistically significant lack of change in a construct over time. If a predictor is more stable, then it is more likely that an initial measurement of this predictor will be relatively consistent over time.

Bandalos (2018, p. 190) illustrates this idea using course placement decisions. If a student takes a placement test to get into a course but the placement test scores are not stable, then that student may be placed in a course that is too difficult or easy for them. In other words, the course placement may not match the student's actual skill level in that domain. This scenario can be applied to non-cognitive constructs. For example, let us imagine that self-efficacy is being measured in high school students to determine which students should be considered for advanced-placement courses. If self-efficacy is not a stable construct, then the findings from the original self-efficacy measurement may not reflect future measurements of that construct. The outcomes of this scenario are like the outcomes presented by Bandalos (2018)-- the advanced-course placement may not match the student's actual skill level in that domain.

Unfortunately, research about the domain specificity and within-person stability of constructs over time is scarce. The aim of this thesis is not only to test which non-cognitive constructs can predict academic performance; this study will also analyze the domain specificity and stability of these traits over time. Analyzing domain specificity and stability over time will help determine appropriate real-life applications for each of these non-cognitive constructs. A construct that is deemed domain specific may not be an appropriate factor to consider in a general academic context, such as the college applications process. Less stable constructs may also not be appropriate to consider when the goal is to foster long-term academic performance.

This study will be analyzing the predictive validity, domain specificity, and stability of the following non-cognitive constructs in relation to secondary and postsecondary academic performance: self-efficacy, utility value, effort cost, beliefs about the nature of intelligence (aka., implicit theories of intelligence), and school engagement. Multiple regression models will be used to analyze the predictive quality of these non-cognitive constructs, and paired t-tests will be conducted to determine these constructs' domain specificity and stability across time.

Data about students' status on these constructs were collected by the National Center for Education Statistics in "a nationally representative, longitudinal study of 9th graders who [were] followed through their secondary and postsecondary years, with an emphasis on understanding students' trajectories from the beginning of high school into postsecondary education, the workforce, and beyond" (NCES, n.d.). This study is called the High School Longitudinal Study of 2009 and will be referred to as HSLS:09 for the remainder of the document. The data collected from the HSLS:09 has been made public by the National Center for Educational Statistics.

Because the HSLS:09 collected data using a longitudinal research design, it provides researchers the opportunity to analyze constructs' development over-time. A longitudinal study allows for the collection of validity evidence since information about within-person differences are collected at multiple timepoints (Bandalos, 2018, p. 288). Finally, HSLS:09 accumulated data about self-efficacy, utility value, and effort cost for mathematics and science domains separately. This will allow us to make comparisons between subjects and assess whether self-efficacy, utility value, and effort cost are significantly different across school subjects (or domains).

In past meta-analyses, some of these motivational predictors have been shown to have positive, significant relationships with academic performance. Schneider and Preckel (2017) reviewed 105 correlates of academic performance from 38 meta-analyses, and self-efficacy and effort regulation were included in this list of correlates. They divided self-efficacy into two separate categories. Performance self-efficacy, defined by the authors as "perceptions of academic performance capability," had the second-strongest association with achievement in higher education (Cohen's d = 1.81). Academic self-efficacy, which is defined as general perceptions of academic capability, was the 21st strongest correlate out of the 105 total correlates analyzed (Cohen's d = 0.58). Effort regulation was the 13th strongest correlate to academic performance (Cohen's d = 0.75).

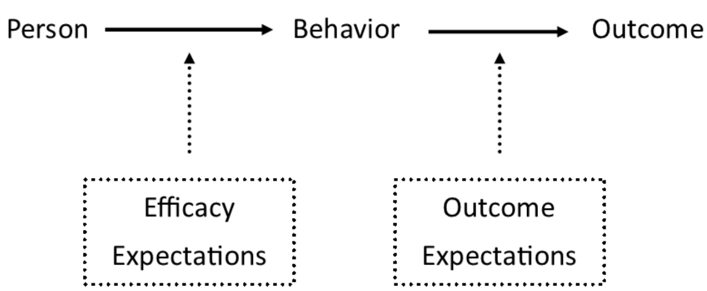
Richardson, Abraham, and Bond (2012) assessed the relationships between the same three non-cognitive constructs and academic performance. Out of the 42 correlates of academic performance assessed by this meta-analysis, performance self-efficacy had the strongest relationship with academic achievement (*r* = 0.59). Academic self-efficacy was the second-strongest motivational correlate of academic performance (*r* = 0.31) and presented a stronger effect size than SAT scores (*r* = 0.29). Effort regulation was categorized as a self-regulatory learning strategy in this study, and it had the strongest relationship with college academic achievement in its category (*r* = 0.32).

These findings demonstrate that non-cognitive constructs can have positive associations with adaptive academic outcomes. Constructs that demonstrate adequate relationships with academic performance indicators can be utilized in academic interventions with the goal of increasing high school and college academic achievement. Or, these constructs (preferably constructs that demonstrate increased construct stability) could be employed in a college admission setting to determine which students are likely to be successful during a postsecondary educational career. Rather than using effect sizes alone to demonstrate relationships between motivational constructs and academic achievement, this thesis utilizes various statistical methods to further explain the relationship between non-cognitive constructs and academic performance as well as further evaluate the potential advantages or disadvantages these constructs possess as predictors of academic success.

LITERATURE REVIEW

Self-Efficacy

Albert Bandura is the researcher primarily associated with the development of the self-efficacy theory. He defines student self-efficacy as "a situation-specific belief that students have on their ability to organize and execute the actions required to learn and master tasks and assignments at a satisfactory level." (Bandura, 1997). Bandura claims that cognitive processes play a prominent role in the acquisition and retention of new behavior patterns, and motivation is also partly rooted in cognitive activities (Bandura, 1977). Self-efficacy theory was developed to further explain these cognitive processes. Self-efficacy theory posits that expectations mediate the relationship between the individual and their behavior as well as the association between one's behavior and eventual outcome (see Figure 1).



Note. Adapted from "Self-efficacy: Toward a unifying theory of behavioral change," by A. Bandura, 1977, *Psychological Review, 84(2), p. 193.* (https://doi.org/10.1037/0033-295X.84.2.191). Copyright 1977 by the American Psychological Association.

Figure 1 *Efficacy Expectations Versus Outcome Expectations*

Two types of expectations are outlined by self-efficacy theory-- efficacy expectations and outcome expectations (Bandura, 1977). As seen in Figure 1, efficacy expectations can affect an individual's behavior. Efficacy expectations are defined as "the conviction that one can successfully execute the behavior required to produce the outcomes" (p. 193). Efficacy expectations determine how much effort people will expend and how long they will persist in the face of obstacles and aversive experiences. Outcome expectations are "a person's estimate that a given behavior will lead to certain outcomes" (p. 193). Although these definitions appear to overlap, Bandura presents a scenario that emphasizes the importance of why these expectations have separate definitions. In this scenario, an individual believes that a specific behavior will produce a specific outcome; however, they don't believe that they can execute the behavior. The individual believing that a specific behavior will produce a specific outcome is the outcome expectation, and the individual not believing that they can personally execute that behavior is a low efficacy expectation. They may understand that adaptive behaviors can lead to positive outcomes, but if their efficacy expectations are low, these expectations are more likely to produce fearful or avoidant behaviors rather than the intended adaptive behaviors.

Three characteristics of efficacy expectations can impact coping behavior (Bandura, 1977). The first characteristic discussed by Bandura is the *strength of conviction*. The individual in the scenario above has a low strength of conviction. They don't believe that they can successfully execute the behavior required to produce the desired outcome. The second characteristic, *magnitude of tasks*, posits that the efficacy expectations of different individuals occur in the simplest tasks as well as more complex tasks. The last characteristic of efficacy expectations is *generality of efficacy*. Some experiences create limited mastery expectations, whereas others instill a more generalized sense of efficacy that extends beyond the specific treatment situation.

To demonstrate the processes that underlie self-efficacy theory, Bandura (1977) conducted a study to understand what types of experiences are more likely to produce adaptive outcomes. Individuals that were adversely afraid of snakes were recruited as participants for this study. The primary aim of this study was to see which of three experiences would allow participants to become most comfortable around snakes. Participants were divided between three experimental conditions: subjects directly engaged in progressively more threatening interactions with a boa constrictor (personal experience), observing another individual engage in progressively more threatening interactions with a boa constrictor (vicarious experience), and no experience. Through this experimental study, Bandura discovered that personal experience tends to affect efficacy expectations the most. This finding highlights the importance of the effect of direct experience on changes in self-efficacy. As participants experienced more successful outcomes over time, their efficacy expectations became more positive. Although vicarious experience also produced an increase in successful performances over time, the increase was not as drastic as individuals that directly engaged in progressively more threatening interactions with the snake.

Self-efficacy is commonly assessed in educational research. In order to clarify the role of self-efficacy in an academic setting, Robbins et al. (2004) provides a definition of academic self-efficacy, which is the self-evaluation of one's ability and/or chances for success in the academic environment. The relationship that connects the individual, their behavior, and an outcome as shown in Figure 1 still stands in an academic context; however, the outcomes in this model are specific to the educational domain.

Academic achievement is a common outcome investigated in research studies. Self-efficacy has been shown in various studies to predict academic performance (Chamarro-Premuzic et al., 2010; Jiang et al., 2018; Olivier et al., 2019; Stajkovic et al., 2018; Villalón et al., 2015). A meta-analysis conducted by Robbins et al. (2004) examined the relationship between psychosocial/study skill factors and college outcomes. Researchers determined that academic self-efficacy was one of the strongest predictors of GPA (r = .496) and retention (r = .359). In fact, self-efficacy was a stronger predictor of college academic performance than both high school GPA and standardized test scores.

Bandura (1997) defines student self-efficacy as a situation-specific construct. In order to confirm this claim, some educational researchers have conducted studies to determine whether students' levels of self-efficacy vary significantly between academic subjects. Research has replicated Bandura's claim regarding the domain specificity of self-efficacy (Bong, 2001; Schöber et al., 2018). Bong (2001) examined between-domain and within-domain relations of self-efficacy in 424 Korean middle and high school students and discovered that self-efficacy was domain specific. Schöber et al. (2018) measured mathematics and reading self-efficacy perceptions for 7th grade students in Germany. The results revealed a low but positive correlation between mathematics self-efficacy and reading-self efficacy at both data collection phases, indicating that self-efficacy was domain specific for this sample.

Longitudinal studies assess whether a variable is stable over time. Although research on the stability of self-efficacy is limited, most studies found domain specific self-efficacy to be stable over the period of a 7th grade academic year (Schöber et al., 2018) and across a 5-year period that collected data from students in their 5th, 7th, and 9th grade years (Petersen & Shibley Hyde, 2015). Mixed results have been presented, however, about the stability of general self-efficacy constructs. Whereas Phan (2009) found general self-efficacy to be moderately stable in undergraduate students (r = .63), Grevenstein and Bluemke's (2017) work has contradicting their findings. They collected data from 15-24-year-olds (N = 299) over a period of 10 years regarding their sense of coherence and general self-efficacy and reported that students' general self-efficacy had low longitudinal stability. Since self-efficacy has been conceptualized as a domain specific construct, it is possible that measuring a version of self-efficacy that does not fit this conceptualization could be the cause of the mixed findings.

Expectancy-Value Theory

Expectancy-value theory was developed by Eccles and her colleagues in 1983 to define the cognitive processes behind predictors of academic performance and choice. The theory states that expectancies, values, and their determinants influence choice, persistence, and performance (Wigfield et al., 2016). Figure 2 presents the Eccles, Wigfield, and colleagues' (2000) expectancy-value model of achievement performance and choice.

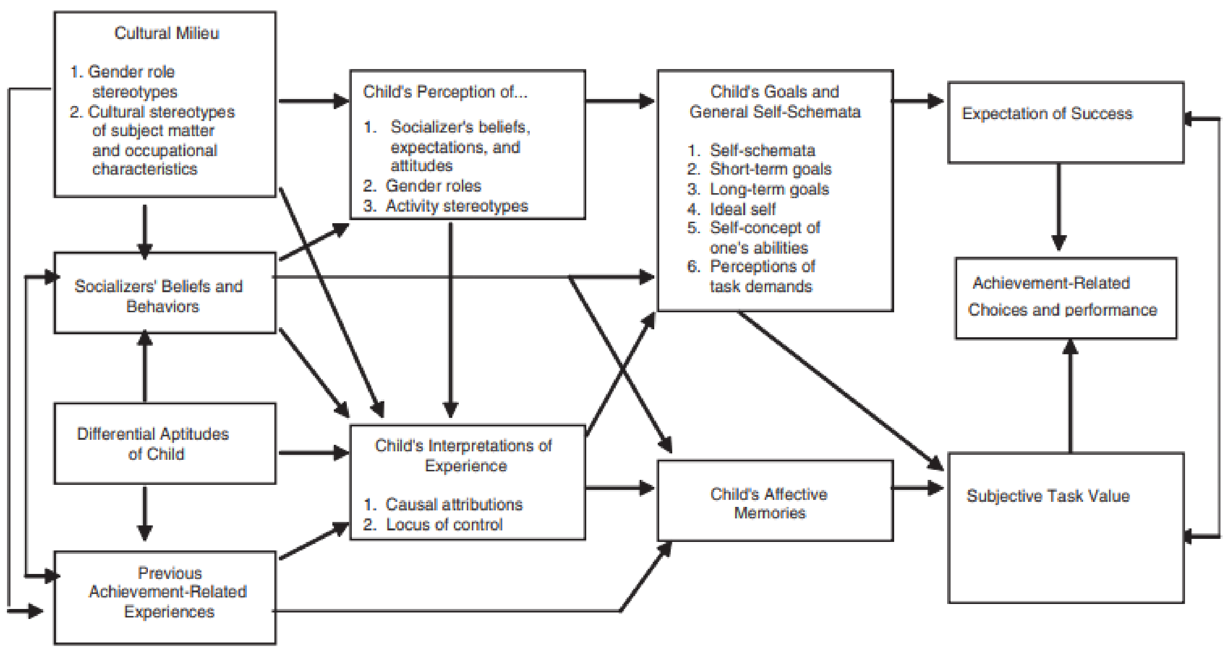


Figure 2 *Expectancy-Value Model*

Note. Adapted from "Expectancy-Value Theory of Achievement Motivation," by A. Wigfield & J. Eccles, 2000, *Contemporary Educational Psychology*, *25*, 68-81. (https://doi.org/10.1006/ceps.1999.1015). Copyright 2009 by Taylor and Francis.

In this model, expectations of success and subjective task values have the most direct relationship to achievement-related choice and performance. Wigfield and Eccles (2000) define expectancy beliefs as one's perception about how well they will do on *upcoming* (immediate or long-term future) tasks. In other words, individuals ask themselves "Am I able to do this upcoming task?" Ability beliefs, which are also commonly referred to as 'competence beliefs', represent an individual's perception regarding their *current* competence at a given activity.

Subjective task values represent the reasons behind someone performing a task in the future. Eccles et al. (1983) proposed four facets of values: *attainment value* (the importance of doing well on a task), *intrinsic value* (the enjoyment gained from doing a task), *utility value* (how a task fits into an individual's future plans), and *cost* (how the decision to engage in one activity limits access to other activities, assessments of how much effort will be taken to accomplish the activity, and its emotional cost). Utility value will be the subjective task value assessed in this thesis.

Wigfield, Tonks, and Klauda (2016) summarize major research findings about how expectancies and subjective task values affect performance and achievement-related choice. Overall, the authors discovered that "students' expectancies for success and beliefs about ability are among the strongest psychological predictors of performance. Students' subjective task values predict both intentions and actual decisions to persist at different activities, such as taking mathematics and English courses and engaging in sports activities." (p. 59). In other words, expectancies have a stronger relationship with performance whereas values have a stronger relationship to achievement-related choice and persistence.

Although value components are often studied in relation to academic choice and persistence, research has been conducted to assess value components' relationship with academic performance. Individuals that believe that a task fits into an individual's future plans (i.e., students with greater perceptions of utility value) are more likely to have higher grades in their courses (Gaspard et al., 2018; Perez et al., 2019; Weidinger et al., 2020). Petersen and Shibley Hyde (2015), however, had contradicting findings in their study. When controlling for prior mathematics performance, neither math utility value nor math interest intercepts or slopes predicted later performance.

Ability beliefs are hypothesized to be a domain-general conceptualization of competence; however, this generalization may not be applicable for subjective task values. Various studies have acknowledged the domain specific nature of the utility value construct. The goal of Gaspard et al. (2018) was to assess the expectancies and value of German students from grades

5-12. They collected data about students' expectancies and values in five academic domains from 857 students and found that utility for job, utility for school, and social utility all had high between-domain correlations. Trautwein et al. (2012) collected expectancy-value data from 2,508 U.S. students at the end of their secondary education careers and utility value was not just domain specific—it was the least domain specific construct evaluated in their study.

The findings regarding the stability of subjective task values over time have been mixed. In a study conducted by Priess-Groben and Shibley Hide (2017), utility value assessed in ninth-grade students was found to be a marginally significant predictor of higher utility value in college. Chouinard and Roy (2008), however, discovered that the utility value of mathematics decreased significantly over the course of two years. A study by Petersen and Shibley Hyde (2015) also supported the findings from Chouinard and Roy (2008); mathematics utility value decreased significantly as students progressed through school. Research on this topic is lacking and contradictory, and one of the aims of this thesis is to produce additional evidence about the stability of subjective task values over time.

Self-efficacy theory and expectancy-value theory have overlapping themes. Bandura's 'efficacy expectation' concept is encapsulated in Eccles, Wigfield, and colleagues' model of expectancy-value theory as an expectation of success. The main distinction between self-efficacy theory and expectancy-value theory lies in the domain specificity of each theory. Self-efficacy is considered a task-specific evaluation of competence (i.e. "Am I confident that I can do this task?") whereas ability beliefs refer to a broader self-evaluation of competence (i.e. "Am I a competent person?"). This distinction is especially important to consider in applied measurement. Assessing both self-efficacy and ability beliefs is not redundant because they evaluated different aspects of competence perception.

Implicit Theories of Intelligence

Dweck and Leggett (1988) developed the implicit theories of intelligence to explain how an individual perceives their own ability. They present two different theories of intelligence: entity and incremental theories. Individuals that hold an *entity* theory of intelligence believe that intelligence is fixed. As a byproduct of this perception, the effort that one exerts will not affect their level of intelligence. Those that possess an *incremental* theory of intelligence believe that intelligence is malleable. Other factors, such as effort exerted, may affect one's intellectual capacity.

The main argument presented by Dweck and Leggett (1988) is that one's theory of intelligence underlies psychological processes that prompts individuals towards one of two specific goal orientations, or reasons for engaging in achievement tasks. Students can either possess a *learning goal orientation,* where the goal is to increase competence through their work, or they can possess a *performance goal orientation*, where the goal is to gain positive judgments and avoid negative judgments of competence. A performance goal orientation does not focus on the increase/decrease of the level of competence; it emphasizes the desire to control others' perceptions of competence. As seen in Table 1, an entity theory of intelligence generally prompts a performance goal orientation, whereas an incremental theory of intelligence often produces a learning goal orientation.

Table 1

Implicit Theories of Intelligence

Theory of Intelligence	Goal Orientation	Perceived Present Ability	Behavior Pattern
Entity (Intelligence is fixed)	*Performance* (Goal: gain positive judgment/avoid negative judgments of competence)	High	*Mastery oriented* (Seek challenge, high persistence)
		Low	*Helplessness* (Avoid challenge, low persistence)
Incremental (Intelligence is malleable)	Learning (Goal: increase competence)	High or low	Mastery oriented

Note: Adapted from "A social-cognitive approach to motivation and personality," by C.S. Dweck & E. L. Leggett, 1988, *Psychological Review*, *95*(2), p. 259. (https://doi.org/10.1037/0033295X.95.2.256). Copyright 1988 by the American Psychological Association.

An individual's perception of their own competence can affect their future behavior. The perception of competence can be viewed on a spectrum from high to low for any individual, regardless of which theory of intelligence they are likely to hold. Dweck and Leggett (1988) present two different behavioral outcome patterns: mastery-oriented behaviors and helplessness. Mastery-oriented behaviors include seeking challenge and maintaining high persistence, and these behaviors are often seen as adaptive. Helplessness is defined as avoiding challenge and having low levels of persistence. The developers of this theory hypothesize that individuals that hold an incremental view of ability are likely to produce mastery-oriented behaviors. These behaviors are likely to occur regardless of how someone views their own competence levels. Varying perceptions in competence, however, does affect future behavioral patterns for individuals that hold an entity theory of intelligence. If someone believes that intelligence is fixed and they perceive themselves as being intelligent, then they are more likely to produce mastery-oriented behaviors. However, for individuals that believe that intelligence is fixed and they perceive themselves as having low intelligence, then they are more likely to exhibit helplessness.

It is hypothesized that individuals who believe that intelligence is malleable, and competence can evolve through exerting effort are likely to show greater academic performance compared to individuals with a fixed mindset about intelligence. Research has confirmed this hypothesis (Blackwell et al., 2007; Costa & Faria, 2018; De Castella & Byrne, 2015; Mouratidis et al., 2017; Romero et al., 2014; Shively & Ryan, 2013); however, associations between incremental views of ability and academic achievement were weak for all studies referenced here. A study conducted by Gunderson et al. (2017) shows that development may be a factor that affects theories of intelligence's relationship with academic achievement. Researchers measured theories of intelligence in four different grade level groups (N = 523): 1st and 2nd grade, 5th and 6th grade, 10th and 11th grade, and college. Correlations between theories of intelligence and self-reported grades were not statistically significant for primary school students; however, high school and college students with more incremental beliefs about math had higher self-reported math grades.

Gunderson et al. (2017) also assessed whether theories of intelligence scores in different domains (reading/writing and math) had statistically significant mean differences. They found that elementary school students did not report statistically significant mean differences in their beliefs about math intelligence versus reading/writing intelligence. However, high school students held more incremental views of reading/writing intelligence than math intelligence. These findings show that theories of intelligence may become more domain specific as students continue through school. A study by Shively and Ryan (2013) corroborated this finding by examining theories of intelligence for college students. The authors found that students' implicit theories of math intelligence were significantly less incremental than their implicit theories of general intelligence. Although these studies present evidence of domain specificity, research on this topic is scarce. I hope to reduce this scarcity by assessing the domain specificity of mathematics and science intelligence perceptions in this thesis.

Only a few studies have assessed the stability of intelligence theories over time. Data collected by Gunderson et al. (2017) showed that mean theories of intelligence scores were significantly different between grade groups, with the highest incremental scores being present in the 5th and 6th grade group for both reading/writing and math. Other studies have presented evidence that theories of intelligence are not stable across time; however, these studies show a different developmental trajectory. A longitudinal study conducted by Romero et al. (2014) shows that middle school students reported more incremental theories of intelligence as they advance from 6th through 8th grade. Shively and Ryan (2013) tracked changes in college students' theories of intelligence over the course of a semester. They found that theories of intelligence became less incremental as the semester progressed. These three studies provide different perspectives about the stability of intelligence theories over time, but there is not enough research currently that replicates their findings.

School Engagement

The most common model of school engagement is a multidimensional, tripartite model that conceptualizes engagement into three different factors: behavioral engagement, emotional engagement, and cognitive engagement. Fredricks et al. (2004) determined that behavioral engagement is defined by a variety of behaviors including adhering to classroom norms, the absence of disruptive behaviors, persistence, concentration, and participation in school-related activities. Emotional engagement refers to students' affective responses in the classroom, such as interest, identification with school, boredom, happiness, sadness, and anxiety. Cognitive engagement emphasizes the psychological investment in learning. Students that are cognitively engaged expend cognitive effort in order to understand the task at hand (Sinatra et al., 2015).

The properties of tripartite school engagement model often intertwine with various motivational and self-regulatory constructs (Sinatra et al., 2015). Effort exerted, for example, is considered a characteristic of behavioral engagement. The belief that exhibiting effort to complete academic tasks is worthwhile has been positively associated with increased academic performance (Perez et al., 2019; Pinxten et al., 2014). Perception of task values (as seen in the

Expectancy-Value theory discussed earlier) can produce emotions that are commonly associated with emotional engagement. Higher levels of self-efficacy are tied to deeper cognitive engagement.

Although the tripartite model of school engagement is the most common operationalization of the school engagement construct, various researchers have presented other operationalizations of school engagement in educational literature. Reeve and Tseng (2011) argue that the addition of an *agentic engagement* dimension better operationalizes school engagement. They define agentic engagement as "students' constructive contribution into the flow of the instruction they receive." (p. 258). Reeve and Tseng (2011) present many examples of constructive contribution, from students offering input or making a suggestion during the flow of instruction to expressing their need for resources or learning opportunities. These actions intentionally involve the student in the instructional process. On the opposite end of the spectrum, Stefansson et al. (2016) argue that a 2-factor model of school engagement, consisting of general school engagement and specific behavioral engagement, best predicts achievement and fits the data as well as the common tripartite model.

Most engagement research shows that students that are more engaged show higher levels of academic achievement (Chase et al., 2014; Dotterer & Lowe, 2011; Lee, 2014; van Rooij et al., 2017). Comparing these studies, however, is difficult because each of the four studies referenced provided different definitions of school engagement. Chase et al. (2014) used the tripartite model as their conceptualization of school engagement and found that all three dimensions significantly predicted self-reported GPA in 710 high school students. They noted, however, that behavioral engagement had the strongest relationship with GPA. Other studies combined dimensions of the tripartite engagement model to answer their research questions. Lee (2014) did not assess cognitive engagement in his sample (N = 3,268 15-year-old students from the United States). This study did conclude that behavioral and emotional engagement significantly predicted reading performance; however, both correlations were weak ($r = .12$ and $r = .09$, respectively). van Rooij et al. (2017) did not assess emotional engagement in his sample and instead analyzed an *intellectual* component of engagement. The researchers compared academic achievement of highly disengaged learners versus engaged learners and they found that higher levels of disengagement in all categories lead to lower levels of academic performance. The main research questions presented by Dotterer and Lowe (2011) was to compare engagement's relationship with academic achievement for struggling and non-struggling learners. They found that engagement formed different relationships with academic achievement for these two groups. All three dimensions of the tripartite model were significantly and positively related with academic achievement in the non-struggling learners group, but behavioral engagement did not have a statistically-significant association with academic achievement in the struggling learners group.

Although the engagement frameworks incorporated into educational research studies is often inconsistent, most of the studies discussed in the paragraph prior assess some form of behavioral engagement. And overall, behavioral engagement was positively associated with various academic performance indicators. In the HSLS:09, behavioral engagement was measured through their school engagement scale as well as through assessment of effort exerted in mathematics and science courses. These behavioral engagement measures will be evaluated in this thesis.

Research Questions and Hypotheses

Similarly to the primary research goal of the National Center for Education Statistics, the primary aim of this thesis is to further understand students' motivational development from the beginning of high school into postsecondary education. Firstly, it is important to determine which motivational constructs predict academic performance. Our hypotheses about the predictive validity of utility value, self-efficacy, and school engagement are based on past literature. Secondly, this thesis aims to assess motivational constructs' domain specificity and stability over time.

Hypothesis 1: Students with higher utility value scores (1a), self-efficacy scores (1b), and school engagement scores (1c) will have greater academic performance (in this study, this includes higher GPAs, higher math achievement assessment scores, and increased likelihood of attending college).

Hypothesis 2: Students with incremental views of ability are more likely to attend college and stay there until graduation.

Hypothesis 3: Utility value (3a) and self-efficacy (3b) are domain specific in line with previous research.

Hypothesis 4: Effort value will be the least domain specific.

<u>Hypothesis 5</u>: Utility value will be the most stable over time.

METHOD

Participant Characteristics and Recruitment

Chapter 3 of the National Center for Education Statistics' Base-Year Data File Documentation outlines the sample design process for the HSLS:09 dataset. Results from an NCES power analysis determined that a minimum of 40 participating public schools per state would be enough to meet the precision criteria set for the national design. To determine anticipated sample sizes for the HSLS:09, NCES researchers conducted two-tailed statistical tests at a 0.05 significance level with the goal of detecting variable mean differences. The researchers explain that two-tailed tests were utilized for two specific reasons: "[they] produce relative standard errors no larger than 2.5 and 10 percent for estimated means and proportions, respectively, within a single wave of the study; and [they] detect a 5 and 15 percentage point change in key estimated means and proportions, respectively, across the study waves" (NCES, 2011, p. 44). A minimum sample size of n = 19,053 would need to be achieved to reach the desired power of 80%.

Researchers utilized a stratified random sampling method to identify eligible schools. Through this process, 1,889 schools in the United States were recruited to take part in the study and a total of 944 of these schools participated in the study. Once study-eligible schools were selected, researchers defined which participants would be eligible to participate in HSLS:09. They determined that an eligible student is any 9th-grade student who attended a study-eligible school in the fall 2009 term. This definition of eligible student is consistent for each data collection phase present in HSLS:09. Out of the total eligible students for each phase of the HSLS:09, only a subset of eligible students completed each item on the student questionnaire or completed both the student questionnaire and the mathematics achievement assessment. The sample sizes for each data collection phase are shown below in Table 2.

Table 2

Sample Sizes by Data Collection Phase

Data Collection Phase	Completed student questionnaire AND math assessment	Completed student questionnaire ONLY
Base Year (BY) *Fall 2009*	*n* = 20,781	*n* = 21,444
First Follow-Up (F1) *Spring 2012*	*n* = 18,507	*n* = 20,594
Undergrad Update (U13) *Summer 2013*	n/a	*n* = 17,656
Second Follow-Up (F2) *Winter 2016*	n/a	*n* = 17,335

Data Collection

Data from the base year and first follow-up student questionnaires and mathematics assessments were collected in-school. Trained assistants conducted the in-school student sessions and would direct participants on how to use school computers or laptop PCs provided by the project to complete the questionnaire and mathematics assessment. Internet was not required for participants to complete the questionnaire or mathematics assessment, and "student responses were stored directly on the laptop in encrypted files and the assistants securely transmitted the data after each in-school session" (NCES, 2011). Two participating schools did not allow in-school sessions. In these instances, student interviews were conducted either through computer-assisted phone interviews or were self-administered online outside of school.

For the U13 data collection phase, the questionnaire could be completed by either the student or parent. Because the U13 questionnaire was 2013 Update was intended to be administered immediately after completion of secondary school, the questionnaire was completed via the web outside of school. The final phase of HSLS:09 consisted of "administering a full-length interview averaging 32 minutes and a 17-minute abbreviated interview through several modes: self-administered via the Web, computer-administered telephone interviewing (CATI), and computer-assisted field interviewing (CAPI)" (NCES, 2018).

Conditions were naturally observed for all phases of the HSLS:09. Manipulation of conditions did not take place in any of the four data collection phases.

Measures

The student questionnaires designed for each data collection phase are extensive and detailed. For the purposes of this thesis, only the measures related to the variables in question will be discussed in detail. Questions used to assess each independent variable are displayed in the appendix for reference. Instrument design was guided using a theoretical model (see Figure 3) that "traces the many influences (including motivation, interests, perceived opportunities, barriers, and costs) on students' values and expectations that factor into their most basic education-related choices" (NCES, 2011, p. 11). Using this model, NCES researchers first identified broad research domains that were considered relevant, and from each domain, key constructs were drawn.

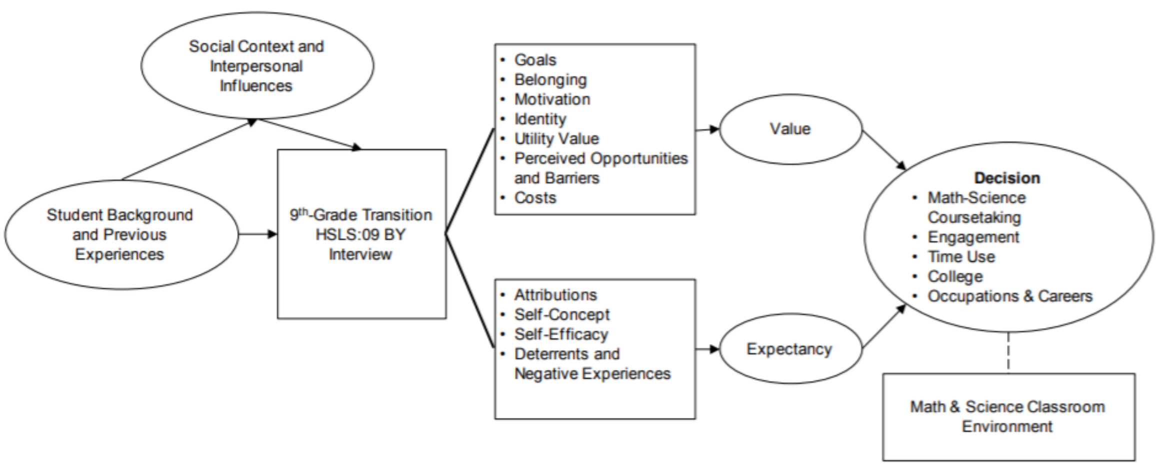


Figure 3 *HSLS:09 Student Survey Conceptual Map*

Note. Adapted from "High School Longitudinal Study of 2009 (HSLS:09) Base-Year Data File Documentation," by The National Center for Education Statistics. (https://nces.ed.gov/surveys/hsls09/usermanuals.asp).

Base-Year Student Questionnaire

This questionnaire collected data from students about demographics, recent school experiences, motivational characteristics in mathematics and science, attitudes about school, math, and science, and future educational and career plans during the fall of 2009. Participants were randomly assigned to one of two groups that determined the order in which sections of the student questionnaire were administered.

Utility Value.

Participants answered six questions related to utility value--three about mathematics utility value and three about science utility value. Students assessed whether mathematics and science were useful in everyday life, useful for college, and useful for a future career. Responses to these questions were presented in a 4-point Likert scale format, where 1 = Strongly agree, 2 = Agree, 3 = Disagree, and 4 = Strongly disagree. Low scores on this scale indicate higher levels of utility value, with the minimum possible score for each domain being a 3 and the highest possible score being a 12.

Self-Efficacy.

Participants answered eight questions related to self-efficacy--four about mathematics self-efficacy and four about science self-efficacy. These self-efficacy items assessed student confidence in subject skill and knowledge mastery. Responses to self-efficacy questions were presented in a 4-point Likert scale format, where 1 = Strongly agree, 2 = Agree, 3 = Disagree, and 4 = Strongly disagree. Low scores on this scale indicate higher levels of self-efficacy, with the minimum possible score for each domain being a 4 and the highest possible score being a 16.

School Engagement.

Participants answered four questions related to school engagement. Students determined how often they exhibited the following behaviors: going to class without homework being completed, going to class without pencil or paper, going to class without books, and going to class late. Responses to these questions were presented in a 4-point Likert scale format, where 1 = Never, 2 = Rarely, 3 = Sometimes, and 4 = Often. Low scores on this scale indicate higher levels of school engagement, with the minimum possible score being a 4 and the highest possible score being a 16.

First Follow-Up Student Questionnaire

This student questionnaire targets the fall 2009 ninth-grade cohort members in the spring term of the 2011-2012 school year regardless of their school enrollment status. Data was collected from this questionnaire about student high school attendance, grade progression and attainment, school experiences, demographics and family background, completion of admissions tests, high school course taking, extracurricular programs, time spent on homework, and jobs/work for pay.

Utility Value.

The questions and scoring process used to assess student mathematics and science utility value is the same as described in the Base-Year Student Questionnaire.

Self-Efficacy.

The questions and scoring process used to assess student mathematics and science self-efficacy is the same as described in the Base-Year Student Questionnaire.

Effort.

Assessment of participant effort exertion was added in the first follow-up student questionnaire. Participants answered eight questions related to effort--four about mathematics effort and four about science effort. These effort items identified features of behavioral engagement (such as paying attention to the teacher and turning assignments in on-time) and cognitive engagement (not giving up on an assignment when difficulties are presented). Responses to these questions were presented in a 5-point Likert scale format, where 1 = Never, 2 = Less than half of the time, 3 = Half of the time, 4 = More than half of the time, and 5 = Always. The minimum possible score for each domain is a 4 and the maximum possible score for each domain is a 20.

Mathematics Assessment in Algebraic Reasoning

The HSLS:09 mathematics assessment was designed to collect data about student algebraic reasoning at two timepoints: 9th and 11th grade. This assessment was administered by computer to participants twice; once during their 9^{th} grade year (BY) and once during their 11^{th} grade year (F1). during the BY and F1 data collection phases. Test scores assessed student understanding of six algebraic content domains (language of algebra; proportional relationships and change; linear equations, inequalities, and functions; nonlinear equations, inequalities, and functions; systems of equations; sequences and recursive relationships) and four algebraic processes (demonstrating algebraic skills; using representations of algebraic ideas; performing algebraic reasoning; solving algebraic problems).

Students answered 40 questions related to algebraic reasoning. All students took a common 15-item Stage 1 test that consisted of 4 items for grade 9 and 11 items for grades 9 and 11. Based on a student's Stage 1 test performance, students were unknowingly directed to a low, moderate, or high level Stage 2 test. Stage 2 was designed so that approximately 25% of students would be routed to the low form, 50% would be routed to the moderate form, and 25% would be routed to the high form. The Stage 2 test consisted of 25 algebraic reasoning questions from the grade 9 and grades 9-11 item banks. The full item bank consisted of 264 unique items. Some items from this test bank were created for grade 9 students, some for both grades 9 and 11, and some items were developed only for grade 11 students.

Item response theory (IRT) was used to develop scores that described student performance on the mathematics achievement assessment. Specifically, a three-parameter logistic (3PL) model was used to estimate students' algebraic reasoning. This model estimates the probability that an individual will correctly respond to an item. Three parameters are utilized in this mathematical model to determine an individual's ability: *a* (discrimination parameter), *b* (item difficulty parameter), and *c* (guessing parameter). An advantage to IRT scoring is that it allows for the comparison of test scores from different test forms. The varied Stage 2 test options make IRT an attractive method for assessing students' algebraic reasoning. A theta value is used to quantify a student's ability level. The higher the theta value is, the higher the ability level. In the context of this mathematics achievement assessment, students with higher theta values show more advanced understanding of algebraic reasoning and concepts.

NCES researchers did not report model fit statistics or estimation procedures for the use of the 3PL model. I did not have access to the raw data necessary to calculate model fit statistics. Model fit statistics are used to ensure that the hypothesized model is the best fit to the data. Future research should request access to this raw data in order to calculate model fit statistics and determine if the 3PL is truly the best-fitting model to the data. Reliability of the mathematics achievement assessment, however, was conducted by NCES researchers by utilizing "a function of the variance of repeated estimates of the IRT ability parameter (within variance), compared with the variability of the sample as a whole." (NCES, p. 32). After sample weights were applied, their IRT-estimated reliability of the HSLS:09 test was 0.92.

Results were discarded if specific indicators concluded that a student was not answering questions to the best of their ability. Missing responses and pattern marking (e.g., all answers were "A" or "ABCDABCDABCD…") prompted discarded results. Results were discarded if there were (1) fewer than six items attempted or (2) pattern marking by selecting the same answer options to more than 10 consecutive items. Less than 1% of math assessment results were discarded using these criteria.

Undergraduate Update Student Questionnaire

For this questionnaire, either a student or a parent could be the respondent. The type of respondent is indicated in the data file. The questionnaire accumulated data about student high school completion, enrollment in courses for college credit, employment, family and military involvement, postsecondary enrollment, and financial aid during the summer of 2013. Independent variables for this thesis will not be collected from this questionnaire; however, total high school GPA values were collected during this data collection phase.

Second Follow-Up Questionnaire

The goal of this student questionnaire was to assess what majors and careers students decide to pursue and the thought processes and experiences behind those decisions. This questionnaire also inquires about students' trajectories from the beginning of high school into postsecondary education, the workforce, and beyond. Data was collected from this questionnaire about student high school experience, postsecondary education, employment, and family/community in the winter of 2016.

Theories of Intelligence.

Participants answered four questions related to theories of intelligence, with two questions assessing perceptions of intelligence in mathematics and two questions assessing perceptions of intelligence in science. To evaluate a student's perception of entity intelligence, participants were instructed to respond to the following question in a 4-point Likert scale format: "You have to be born with the ability to be good at [math/science]," where 1 = Strongly agree, 2 = Agree, 3 = Disagree, and 4 = Strongly disagree. The same response scale was presented for the following items about incremental intelligence: "Most people can learn to be good at [math/science]." A low score for each question represents a stronger affiliation with the theory of intelligence in question.

Academic Performance Variables

High school GPA (HSGPA).

Grade point average (or GPA) is calculated by the National Center for Education Statistics (2011) by "dividing a student's total grade points earned by the total course credits attempted." The most common GPA range is a 4-point scale ranging from 0.0 - 4.0, where a student with a 0.0 GPA has been awarded F's in all of their courses and a student with a 4.0 GPA has been awarded A's in all of their courses. Higher GPAs indicate increased academic performance. The range of GPAs observed in the HSLS:09 data is 0.25 to 4.0. High school GPAs for all grade levels (9th, 10th, 11th, and 12th) were collected for the HSLS:09 as well as a cumulative high school GPA value (HSGPA). HSGPA will be the only GPA indicator utilized in data analysis for this thesis.

Mathematics assessment scores.

In the context of the Mathematics Assessment in Algebraic Reasoning, students with higher theta values show more advanced understanding of algebraic reasoning and concepts.

College enrollment.

In the fourth phase of data collection for the HSLS:09, researchers collected data regarding individuals' current college enrollment status. All participants had to respond "Yes" or "No" to the following question: "Did you attend any college or trade school between the time you [received your high school diploma/received your GED/received your high school equivalency/received your certificate of attendance or completion/last attended high school] and February 2016?" If a participant answers "No" to this question, this means that the participant has not attended any college or trade/technical schools since their last attendance in high school.

Data Analysis

Many of the variables analyzed in this thesis are composite variables, which are "generated using responses from two or more questionnaire items or from recoding of a variable (typically for disclosure avoidance reasons)" (NCES, p. 175). These composite variables have undergone imputation procedures, making them better estimates in data analysis. The NCES claims that they can be used as either classification variables or independent variables in analysis. In order to improve estimation, all variables analyzed in this thesis will be composite variables apart from variables about theories of intelligence. Questions about intelligence theories were not combined in the HSLS:09 data set to create an overall intelligence theory scale score.

Questionnaire composite scores were reverse coded to equate larger scale values with strong associations with the attribute. For example, a higher composite value for math self-efficacy indicates higher levels of math self-efficacy. This framework applies to all motivational constructs assessed during the first and second phases of data collection. All motivational composite variables were created through principal components factor analysis (weighted by a student sampling weight) and standardized to a mean of 0 and standard deviation of 1. Only respondents who provided a full set of responses were assigned a scale value.

Variables associated with theories of intelligence were not transformed into composite variables. In the second follow-up questionnaire, lower scores on each question represented stronger affiliations with the intelligence theory in question. In order to keep the analytic framework consistent, items regarding theories of intelligence were reverse coded so that higher values represented stronger affiliations with the related theory of intelligence.

Various analytic methods will be employed in this thesis. To assess preliminary relationships between variables, point-biserial correlations will be computed. To analyze construct domain specificity and construct stability, paired t-tests will be conducted with the purpose of determining whether mean differences between two constructs are statistically significant. Finally, to assess the predictive nature of independent variables, regression models will be employed. For our continuous outcomes, HSGPA and mathematics achievement, multiple linear regression models will be used. College enrollment will be treated as a dichotomous outcome, so a logistic regression model will be created to evaluate relationships between predictors and predicted college enrollment. A full list of variables used in analyses is presented below in Table 3.

Table 3

HSLS:09 Full Variable List

Data Collection Phase	Variable Name	Variable Code	Predictor (P) or Outcome (O)
Base Year	Math Utility	X1MTHUTI	P
	Math Efficacy	X1MTHEFF	P
	Science Utility	X1SCIUTI	P
	Science Efficacy	X1SCIEFF	P
	Engagement	X1SCHOOLENG	P
	Math Achievement	X1TXMTH	O*
First Follow-Up	Math Utility	X2MTHUTI	P
	Math Efficacy	X2MTHEFF	P
	Math Effort	X2MEFFORT	P
	Science Utility	X2SCIUTI	P
	Science Efficacy	X2SCIEFF	P
	Science Effort	X2SEFFORT	P
	Math Achievement	X2TXMTH	O
U13 Update	HSGPA	X3TGPATOT	O
Second Follow-Up	Math Incremental TOI	S4MLEARN	P
	Sci Incremental TOI	S4SLEARN	P
	College Enrollment	X4EVRATNDCLG	O

Note. Variable codes starting in "X" represent composite variables. Variables starting with "S" indicate variables with raw item scores. O* was included in correlation analyses but not in regression analyses.

RESULTS

Demographic information was only collected during the BY and F1 data collection phases. Tables 4 and 5 present unweighted demographic characteristics for these two data collection periods. Sex was approximately equally distributed for both data sets. The majority of participants identified as white (approximately 55% of BY sample and 56% of F1 sample) and middle class (approximately 58% of BY sample and 58% of F1 sample). Most schools represented in both samples were public high schools located in Southern suburbia. These schools hailed on the urban fringe of large or mid-size cities in states colored blue in Figure 4.

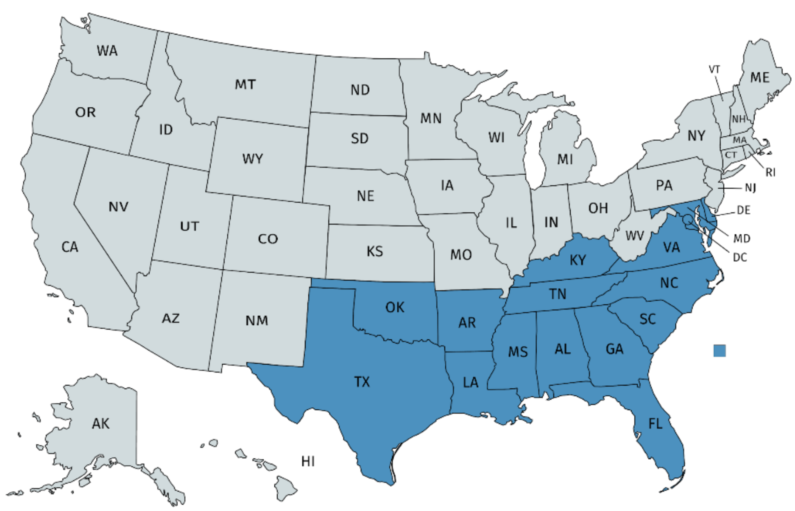


Figure 4 *South Region of HSLS:09 Data Collection*

Table 4

HSLS:09 Unweighted Demographic Characteristics (Base Year)

	Completed student survey		Completed survey + math assessment	
	n	%	*n*	%
Total	21,444		20,781	
Sex				
Male	10,887	50.77%	10,529	50.67%
Female	10,557	49.23%	10,252	49.33%
Missing value		0.00%	0	0.00%
Race/Ethnicity				
American Indian/Alaska Native	163	0.76%	155	0.75%
Asian	1,672	7.80%	1,621	7.80%
Black or African American	2,218	10.34%	2,151	10.35%
Hispanic	3,515	16.39%	3,395	16.34%
Native Hawaiian/Pacific Islander	110	0.51%	106	0.51%
White	11,854	55.28%	11,518	55.43%
More than one race	1,912	8.92%	1,835	8.83%
Missing value		0.00%		0.00%
SES				
Low SES	3,434	16.01%	3,284	15.80%
Mid SES	12,491	58.25%	12,149	58.46%
High SES	5,519	25.74%	5,348	25.74%
Region				
Northeast	3,331	15.53%	3,217	15.48%
Midwest	5,695	26.56%	5,508	26.50%
South	8,705	40.59%	8,471	40.76%
West	3,713	17.31%	3,585	17.25%
Locale				
City	6,067	28.29%	5,886	28.32%
Suburban	7,636	35.61%	7,372	35.47%
Town	2,580	12.03%	2,512	12.09%
Rural	5,161	24.07%	5,011	24.11%
School Type				
Public	17,511	81.66%	16,928	81.46%
Catholic or other private	3,933	18.34%	3,853	18.54%

Note. Categories for socioeconomic status (SES) were defined using the SES quintile variable (X1SESQ5), where X1SESQ5 = 1 (20^{th} percentile) represents low SES and X1SESQ5 = 5 (80^{th} percentile) represents high SES. All others were classified as middle SES.

Table 5

HSLS:09 Unweighted Demographic Characteristics (First Follow-up)

	Completed student survey		Completed survey + math assessment	
	n	%	*n*	%
Total	20,594		18,507	
Sex				
Male	10,384	50.42%	9,266	50.07%
Female	10,210	49.58%	9,241	49.93%
Race/Ethnicity				
American Indian/Alaska Native	142	0.69%	123	0.66%
Asian	1,675	8.13%	1,580	8.54%
Black or African American	2,121	10.30%	1,817	9.82%
Hispanic	3,271	15.88%	2,873	15.52%
Native Hawaiian/Pacific Islander	97	0.47%	85	0.46%
White	11,532	56.00%	10,475	56.60%
More than one race	1,756	8.53%	1,554	8.40%
SES				
Low SES	3,167	15.38%	2,736	14.78%
Mid SES	12,066	58.59%	10,820	58.46%
High SES	5,361	26.03%	4,951	26.75%
Region				
Northeast	3,169	15.39%	2,911	15.73%
Midwest	5,346	25.96%	4,882	26.38%
South	8,261	40.11%	7,448	40.24%
West	3,350	16.27%	2,960	15.99%
Component not applicable	236	1.15%	145	0.78%
Missing value	232	1.13%	161	0.87%
Locale				
City	5,629	27.33%	5,061	27.35%
Suburban	6,146	29.84%	5,526	29.86%
Town	2,598	12.62%	2,364	12.77%
Rural	5,756	27.95%	5,251	28.37%
Component not applicable	236	1.15%	145	0.78%
Missing value	229	1.11%	160	0.86%
School Type				
Public	16,797	81.56%	15,089	81.53%
Catholic or other private	3,336	16.20%	3,115	16.83%
Component not applicable	236	1.15%	145	0.78%
Missing value	225	1.09%	158	0.85%

Note. Categories for socioeconomic status (SES) were defined using the SES quintile variable (X1SESQ5), where X1SESQ5 = 1 (20^{th} percentile) represents low SES and X1SESQ5 = 5 (80^{th} percentile) represents high SES. All others were classified as middle SES. "Component not applicable" is filled for all variables across the entire questionnaire when a component did not apply (e.g., parents not included in the F1 subsample).

Summary statistics were calculated for measures used to assess school engagement as well as mathematics and science utility value, self-efficacy, effort cost, and theories of intelligence (see Tables 6-8). Raw total scale scores were used to produce summary information for this thesis. A lower total score equates to a stronger association with that variable. Cronbach's α for all BY and F1 student scales ranged from 0.74-0.92. According to George and Mallery (2003), higher values indicate greater reliability, and all scales in this thesis have acceptable reliability coefficients (i.e. $\alpha \geq .70$). Cronbach's α was not calculated for questions about theories of intelligence. Researchers formulated these questions with the intention of them individually assessing student incremental/entity intelligence theory attachment (rather than combining these questions to make an overall intelligence theory score).

Table 6

Summary Information for Raw Student Motivation Variables (Base Year)

Student scales	Mean	SD	Range	Cronbach's α
Mathematics utility	5.55	1.87	(3, 12)	0.78
Mathematics self-efficacy	8.24	2.64	(4, 16)	0.90
School engagement	7.39	2.44	(4, 16)	0.68
Science utility	6.25	1.86	(3, 12)	0.75
Science self-efficacy	8.62	2.53	(4, 16)	0.88

Note. The statistics provided for these scales are derived from the sample that completed both the student questionnaire and the mathematics assessment (n = 20,781)

Table 7

Summary Information for Raw Student Motivation Variables (First Follow-Up)

Student scales	Mean	SD	Range	Cronbach's α
Mathematics utility	5.14	1.77	(3, 12)	0.82
Mathematics self-efficacy	8.88	2.85	(4, 16)	0.89
Mathematics effort (raw)	12.2	1.66	(4, 20)	0.74
Mathematics effort (reverse)	4.2	0.71		0.74
Science utility	5.81	1.94	(3, 12)	0.82
Science self-efficacy	8.73	2.92	(4, 16)	0.92
Science effort (raw)	11.99	1.67	(4, 20)	0.75
Science effort (reverse)	4.2	0.72		0.75

Note. The statistics provided for these scales are derived from the sample that completed both the student questionnaire and the mathematics assessment (n = 18,507).

Table 8

Summary Information for Theories of Intelligence (Second Follow-Up)

Student scales	Mean	SD	Range
Entity theory (mathematics)	2.86	0.76	(1, 4)
Entity theory (science)	2.91	0.71	(1, 4)
Incremental theory (mathematics)	1.99	0.65	(1, 4)
Incremental theory (science)	1.95	0.59	(1, 4)

Based on the summary information, it can be concluded that participants seem to possess stronger utility value for mathematics than science (Base Year: $\bar{x}_m = 5.55$, $\bar{x}_s = 6.25$; First Follow-Up: $\bar{x}_m = 5.14$, $\bar{x}_s = 5.81$). Although participants believed that mathematics was more useful in everyday life, for college, and for a future career throughout all high school data collection phases, they were less consistent about their perceptions of self-efficacy from their base year to first follow-up student questionnaires. In the 9th grade data collection phase, participants reported feeling more confident about their ability to succeed in mathematics courses than science courses ($\bar{x}_m = 8.24$, $\bar{x}_s = 8.62$). However, at the end of their 11th grade academic year, students believed that they were more capable of learning and mastering tasks and assignments at a satisfactory level in science than mathematics courses ($\bar{x}_m = 8.88$, $\bar{x}_s = 8.73$).

In order to compare descriptive statistics for math effort and science effort variables, Questions 3 and 4 of their assessments were reverse coded so that lower scale scores represented more effort exerted in that school subject. Allowing lower scores to represent increased effort exertion aligns with the interpretation framework of the other motivational characteristics analyzed in this thesis. These transformed variables are exhibited in Table 7 using (reverse) as notation. Participant self-reports of effort exerted in their mathematics and science courses revealed comparable effort exertion between domains (Math: $\bar{x} = 4.2$ and $s = 0.71$; Science: $\bar{x} = 4.2$ and $s = 0.72$).

These results also suggest that participants perceived greater utility value for both mathematics and science as students progress through their high school education (Math: $\bar{x}_{BY} = 5.55$, $\bar{x}_{F1} = 5.14$; Science: $\bar{x}_{BY} = 6.25$, $\bar{x}_{F1} = 5.81$). The opposite was true, however, for student self-efficacy. As participants continued through their high school years, they became less confident in their abilities to master tasks at a satisfactory level in both mathematics and science courses (Math: $\bar{x}_{BY} = 8.24$, $\bar{x}_{F1} = 8.88$; Science: $\bar{x}_{BY} = 8.62$, $\bar{x}_{F1} = 8.73$).

Summary information was calculated for both administrations of the math achievement assessment. The average theta score of participants was larger in the second administration of the math assessment (θ = 0.55, s = 1.13) than in the initial administration of the math achievement assessment (θ = -0.07, s = 0.97). Overall, students exhibited more advanced reasoning of algebraic concepts in their 11th grade year compared to their 9th grade fall semester. The minimum theta value present in both administrations of the assessment was θ = -2.60; however, the maximum theta values in both administrations are different, with the baseline assessment maximum, θ = 3.00, being smaller than the follow-up assessment maximum theta value of θ = 4.50.

The distribution of theta values for both administrations are illustrated in Figures 5 and 6. Average mean theta values are indicated in these figures by a red line. The distribution of theta values for both administrations appear slightly left-skewed; however, the theta value distribution of the initial assessment appears to be unimodal near the mean theta value whereas the distribution of theta values for the follow-up math achievement assessment are bimodal. The two modes for the follow-up math achievement distribution are around the mean theta value and around θ = 2.50.

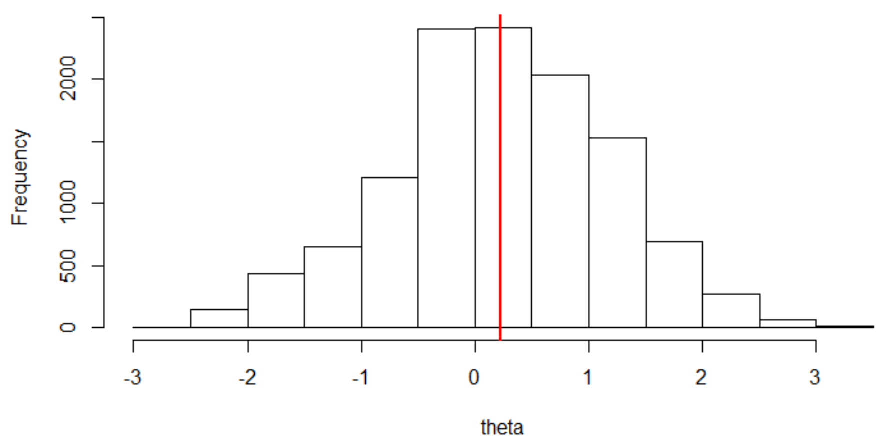


Figure 5 *Distribution of Mathematics Achievement Assessment Theta Values (Base Year)*

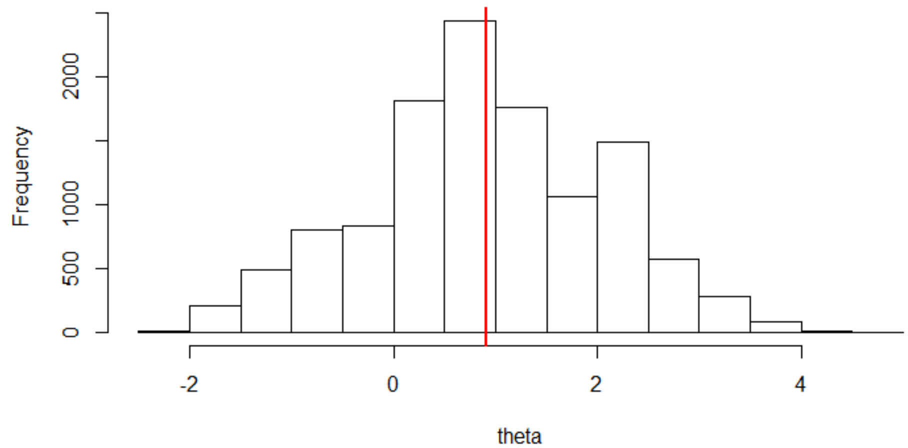


Figure 6 *Distribution of Mathematics Achievement Assessment Theta Values (First Follow-Up)*

For the remaining data analyses, it was important to ensure that participants answered all questions about the variables of interest in this thesis. If a participant did not respond to all questions of interest across all data collection phases, their data was completely removed from the data set. By only preserving entries that completely responded to the targeted questions for all data collection phases, this further ensures that the appropriate information required to make longitudinal conclusions is present. A sample of N = 5,789 was used for the remainder of data analysis for this thesis. RStudio was utilized for all data analyses in this thesis (RStudio Team, 2015). The following RStudio packages were installed to assist in analyses: tidyverse, psych, lavaan, ggplot2, and leaps. Demographic characteristics for this final analytic sample are presented in Table 9.

Table 9

HSLS:09 Unweighted Demographic Characteristics (Final Analytic Sample)

	n	%
Total	5,789	
Sex		
Male	2,687	46.42%
Female	3,102	53.58%
Race/Ethnicity		
American Indian/Alaska Native	29	0.50%
Asian	605	10.45%
Black or African American	439	7.58%
Hispanic	766	13.23%
Native Hawaiian/Pacific Islander	26	0.45%
White	3,455	59.68%
More than one race	469	8.10%
Missing value	0	0.00%
SES		
Low SES	552	9.54%
Mid SES	3,031	52.36%
High SES	2,206	38.11%
Region		
Northeast	1,033	17.84%
Midwest	1,804	31.16%
South	2,042	35.27%
West	910	15.72%
Locale		
City	1,797	31.04%
Suburban	2,211	38.19%
Town	579	10.00%
Rural	1,202	20.76%
School Type		
Public	4,329	74.78%
Catholic or other private	1,460	25.22%

Compared to the original demographic composition of HSLS:09 (refer to Tables 4 and 5), the final analytic sample has more female respondents than male respondents. The majority of participants identified as white and middle class, which follows the demographic information of the full sample. There was an increased percentage of high SES participants in the final sample when compared with the original demographic data. Just as is presented in the original demographic findings, most schools represented in the final analytic sample were public high schools located in Southern suburbia. The final analytic sample had a higher percentage of private school respondents compared to the BY and F1 full samples.

In order to determine benchmarks for correlation strengths, we turn to Bosco et al.'s (2015) large-scale analysis of 147,328 correlational effect sizes (*r*s) in applied psychology research. As part of this study, researchers categorized variables observed in applied psychology academic journal articles (examples include attitudes, intentions, behavior, and performance). Most of the variables used in this thesis would be classified as attitude variables. The authors found that $.18 \leq |r| < .39$ was considered a "moderate" correlational size between two attitude variables. Correlational effect sizes above $|r| = .39$ exceed the 67th effect size distribution (ESD) percentile, and any effect size included in the range $-.18 < r < .18$ would not reach the 33rd percentile. This will be the framework that will be utilized in this thesis when interpreting correlational effect sizes.

A Pearson correlation matrix (see Table 10) was developed to evaluate associations between motivational constructs and the two continuous academic performance indicators (high school GPA and mathematics achievement). The two strongest correlations present were between BY and F1 math achievement assessment scores ($r = .76$) and between math and science incremental intelligence beliefs ($r = .65$). Ability scores from the base year administration of the mathematics achievement assessment are highly associated with ability scores from the second administration of the assessment (and vice versa).

Table 10

Pearson Correlation Matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Math Utility (BY)	1	1.00																		
Math Efficacy (BY)	2	.35	1.00																	
Sci Utility (BY)	3	.43	.20	1.00																
Sci Efficacy (BY)	4	.18	.39	.38	1.00															
Engagement (BY)	5	.09	.19	.14	.19	1.00														
Math Achieve (BY)	6	.01	.30	.09	.27	.16	1.00													
Math Utility (F1)	7	.32	.22	.20	.12	.10	.14	1.00												
Math Efficacy (F1)	8	.18	.36	.14	.26	.12	.25	.40	1.00											
Sci Utility (F1)	9	.17	.16	.33	.24	.10	.19	.45	.24	1.00										
Sci Efficacy (F1)	10	.11	.23	.13	.29	.06	.16	.19	.29	.37	1.00									
Math Effort (F1)	11	.12	.18	.15	.15	.27	.14	.28	.44	.18	.14	1.00								
Sci Effort (F1)	12	.09	.14	.15	.17	.21	.11	.16	.18	.30	.47	.50	1.00							
Math Achieve (F1)	13	.01	.29	.09	.27	.16	.76	.18	.30	.21	.19	.20	.14	1.00						
HSGPA	14	-.03	.20	.07	.21	.29	.52	.13	.20	.18	.11	.33	.26	.56	1.00					
Math Increment (F2)	15	.09	.13	.07	.10	.00	.06	.15	.18	.09	.08	.07	.02	.09	.00	1.00				
Math Entity (F2)	16	-.04	-.04	-.03	-.04	.02	.01	-.08	-.09	-.04	-.03	-.03	-.01	.00	.03	-.35	1.00			
Sci Increment (F2)	17	.06	.09	.09	.13	-.02	.05	.11	.12	.13	.12	.04	.06	.07	.00	.65	-.26	1.00		
Sci Entity (F2)	18	-.02	-.02	-.03	-.03	.02	.01	-.06	-.07	-.04	-.04	-.02	-.01	.00	.02	-.25	.77	-.36	1.00	
College Enrollment (F2)	19	-.03	.09	.03	.10	.13	.27	.05	.11	.09	.07	.18	.14	.31	.41	-.02	.05	.01	0.02	1.00

These correlations (as well as the regression analyses discussed later in this chapter) can be used to assess relationships between the same motivational constructs in different subject matters. Moderate to strong positive correlations were present between math utility value and science utility value (r_{BY} = .38, r_{F1} = .45). This association, however, became stronger over time. Utility value perceptions were only collected during two of the data collection phases of the HSLS:09; however, collecting utility value information from the last two data collection phases would help inform whether utility value becomes less domain specific over time. The associations between math self-efficacy and science self-efficacy were also moderate and positive for both timepoints (r_{BY} = .39, r_{F1} = .29). The motivational variables with the strongest across-domain correlation were the math and science effort variables (r = .50). In other words, effort exerted in high school courses is the least domain specific variable present in this thesis. Additional data analysis will be conducted to further analyze this claim.

Stability over time can also be analyzed using Pearson correlations. Mathematics self-efficacy had the strongest correlation between timepoints (r = .36); however, science self-efficacy had the weakest correlation between BY and F1 (r = .29). This finding illuminates the domain specific nature of self-efficacy, but additional data analysis will be conducted to determine the statistical significance of this claim. Utility values constructs were the only other independent variables that can be compared over time. Moderate, positive correlations connect the two utility value variables over-time (r_M = .32, r_S = .33).

There is a strong, positive association between mathematics and science incremental intelligence beliefs (r = .65). In other words, students that believe that math intelligence is malleable may also be likely to believe that science intelligence is malleable. This finding may indicate that intelligence beliefs are not domain specific, but future data analysis will test this claim further. The second-strongest association with mathematics incremental belief is F1 math self-efficacy (r = .18). BY math self-efficacy had an even weaker association with malleable intelligence beliefs regarding the math domain (r = .13). School engagement had weak or negligible associations with math and science incremental beliefs (r_M = .00, r_S = -.02). HSGPA had negligible relationships with both incremental beliefs as well (r_M = .00, r_S = .00).

Out of all motivational characteristics included in the correlation matrix, effort exerted in high school math courses had the strongest positive association with HSGPA (r = .33), followed closely by school engagement (r = .29) and science effort (r = .26). Math and science effort variables did not have the strongest relationships with mathematics achievement scores (r_M = .20, r_S = .14). For the initial administration of the math assessment, BY math self-efficacy had the strongest relationship with algebraic understanding (r = .30). This trend was also seen for the follow-up administration of the mathematics assessment; F1 math self-efficacy had a moderate association with F1 math achievement scores (r = .30). Science self-efficacy's association with algebraic reasoning became weaker over time (r_{BY} = .27, r_{F1} = .19). BY and F1 science utility value had stronger relationships with algebraic reasoning scores (r = .09-.21) than mathematics utility value (r = .01-.18) at both administrations of the mathematics assessment.

Out of all variables correlated with college enrollment, HSGPA (r = .41) and both mathematics achievement assessment scores (BY: r = .27, F1: r = .31) had the strongest, positive relationships with college enrollment. BY math utility value had a negative, but negligible correlation with future college enrollment (r = -.03). Although math utility value had a positive correlation with college enrollment after some time had passed, this correlation was still weak (r = .05). Science utility value had stronger relationships with college enrollment when compared to math utility value (BY: r = .03, F1: r = .09), but these values are still weak overall. Therefore, it can be concluded that utility value has a minimal relationship to future college enrollment. This thesis hypothesized that students with higher utility value, self-efficacy, and school engagement scores are more likely to attend college. Hypothesis 1 is not supported by the point bi-serial correlation results.

Self-efficacy had stronger relationships with college enrollment compared to utility value, but self-efficacy correlation values for both math and science domains only range from r = .07- .11. The hypothesized relationships between self-efficacy and college enrollment are only minutely supported by these results. This is also the case with the hypothesized relationship between school engagement and college enrollment (r = .13). The only independent variable not included in Hypothesis 1, effort exertion, ironically had the strongest relationships with college enrollment (Math: r = .18, Science: r = .14).

Hypothesis 2 theorized that students that possess an incremental view of intelligence are more likely to attend college. This hypothesis is not supported by the point bi-serial correlation results. The relationship between a malleable view of mathematics intelligence and college enrollment was not significantly different from zero (r = -.02, $p > 0.05$). Although the correlation between incremental views of science intelligence and college enrollment was positive, the correlation itself is also negligible (r = .01, p > 0.05). These values allude to a minimal relationship between holding an incremental view of intelligence and increased chances of college enrollment.

In order to determine the domain specificity of motivational constructs, paired t-tests were conducted for the same construct of varying domains. Table 11 presents paired t-test results related to testing for domain specificity. These results suggest that utility value, BY self-efficacy, and incremental beliefs are domain specific. Effect sizes for all domain specific pairs, however, were deemed small by Cohen's (1988) standards, where d = 0.2 is considered a small effect size. The pair with the large Cohen's d value was the math/science entity belief pair (d = 0.12). The means differ by 0.12 the standard deviation of the data. It can be concluded by these small effect sizes that the statistical significance of the paired t-test results could have been influenced by the large sample size.

Only two paired t-tests conducted in analysis did not show a significant difference between means. Math effort and science effort were not considered domain specific (t = 1.79, p = 0.07), which supports Hypothesis 4. This outcome is further supported by the comparable mean values of effort exerted between domains discussed prior. Additionally, math self-efficacy and science self-efficacy were not domain specific during the second data collection phase (t = 1.81, p = 0.07). This finding about self-efficacy contradicts results calculated for the initial data collection phase.

Table 11

Paired T-Test Results: Domain Specificity

Pairs	Mean	SD	*t*(5788)	*p*	*d*
BY Math utility	5.55	1.87	-4.57	< 0.001	0.06
BY Science utility	6.25	1.86			
BY Math self-efficacy	8.24	2.64	4.00	< 0.001	0.05
BY Science self-efficacy	8.62	2.53			
F1 Math utility	5.14	1.77	5.27	< 0.001	0.07
F1 Science utility	5.81	1.94			
F1 Math self-efficacy	8.88	2.85	1.81	0.07	0.02
F1 Science self-efficacy	8.73	2.92			
Math effort	12.2	1.66	1.79	0.07	0.02
Science effort	11.99	1.67			
Math entity belief	2.86	0.76	8.94	< 0.001	0.12
Science entity belief	2.91	0.71			
Math incremental belief	1.99	0.65	-6.78	< 0.001	0.09
Science incremental belief	1.95	0.59			

Paired t-tests were also utilized to determine constructs' stability over time (see Table 12 for t-test statistics). Math utility value, science utility value, math self-efficacy, and science self-efficacy were the only independent variables of interest in this thesis assessed at multiple timepoints during the HSLS:09 data collection process. Paired t-test results with a p-value > 0.05 do not have a statistically significant difference between means, hence these variables would be considered stable over-time.

Hypothesis 5 was not supported by paired t-test results; utility value was not the most stable over time. Science self-efficacy was the only construct determined to be stable over time (t = 1.19, p = 0.24). This finding shows that an individual's level of science self-efficacy in the science field stays consistent and is unlikely to waver. If science self-efficacy is shown to be a viable predictor of academic performance, then the stability of science self-efficacy may further support the notion that science self-efficacy could be used as a long-term predictor of academic achievement. Although science self-efficacy was considered stable over the course of two years, math self-efficacy, was not considered stable (t = 3.11, p = 0.002). The effect size for the science self-efficacy pair, however, is almost negligible (d = 0.02). As discussed for the domain specificity paired t-test results, it is possible that the statistical significance of the science self-efficacy paired t-test may have been impacted by the large sample size.

Table 12

Paired T-Test Results: Construct Stability

Pairs	Mean	SD	*t*(5788)	*p*	*d*
BY Math utility	5.55	1.87	-4.22	< 0.001	0.06
F1 Math utility	5.14	1.77			
BY Science utility	6.25	1.86	5.24	< 0.001	0.07
F1 Science utility	5.81	1.94			
BY Math self-efficacy	8.24	2.64	3.11	0.002	0.04
F1 Math self-efficacy	8.88	2.85			
BY Science self-efficacy	8.62	2.53	1.19	0.24	0.02
F1 Science self-efficacy	8.73	2.92			

Regression models were used in this thesis to assess motivational constructs' ability to predict academic performance indicators. Because cumulative HSGPA and mathematics achievement assessment scores are continuous, dependent variables, multiple linear regression models were used to determine predictive ability. The dependent variable for college enrollment, however, is categorical; therefore, a linear regression model is not an appropriate model structure for this variable. A logistic regression model was used to analyze associations between independent variables and college enrollment. This model presents relationships between predictors and college enrollment in a S-shaped curve, where 0 denotes a student not being enrolled in college in 2016 and where 1 indicates that a student was enrolled in at least one college course in 2016.

The intelligence belief constructs are conceptualized in a categorical fashion; however, data on entity and intelligence beliefs were collected in such a way that the variables are ordinal. We can treat these variables as approximately continuous in the regression models, where a higher value represents a stronger association with the theory in question. Since the entity and incremental items assess opposing views of intelligence perception, only the assessments of science and mathematics incremental beliefs will be included in the regression models to prevent redundancy.

Table 13 displays the full model that each regression started with and compares this full model to each of the model's reduced regression. The goal of reducing a regression model is to reduce multicollinearity and determine which independent variables statistically impact the model (Montgomery, Peck, & Vining, 2012). Variable selection techniques are used to systematically remove nonsignificant variables from a regression model. Backwards elimination was used to reduce each model. In each step of the elimination process, the variable with the largest *p*-value was removed until all independent variables had a *p*-value < 0.05.

Table 13

Comparisons of Full Regression Model with Reduced Regressions

Full Model	Reduced Model: HSGPA	Reduced Model: Math Assessment	Reduced Model: College Enrollment
BY Math utility	BY Math utility	BY Math utility	BY Math utility
BY Math efficacy	BY Math efficacy		BY Math efficacy
BY Science utility		BY Science utility	
BY Science efficacy	BY Science efficacy	BY Science efficacy	BY Science efficacy
School engagement	School engagement	School engagement	School engagement
F1 Math utility		F1 Math utility	
F1 Math efficacy	F1 Math efficacy		F1 Math efficacy
F1 Math effort	F1 Math effort	F1 Math effort	F1 Math effort
F1 Science utility	F1 Science utility	F1 Science utility	F1 Science utility
F1 Science efficacy	F1 Science efficacy	F1 Science efficacy	
F1 Science effort	F1 Science effort	F1 Science effort	F1 Science effort
Math incremental belief	Math incremental belief	Math incremental belief	Math incremental belief
Science incremental belief			

After using backwards elimination to remove nonsignificant variables from each linear regression model, it was shown that the models explained 20.63% of the variation in HSGPA and 14.14% of the variation in math achievement assessment scores. The independent variables included in the reduced multiple linear regression models were determined to be significant predictors of HSGPA [F (10, 5778) = 150.2, p = < 0.001] and math achievement assessment scores [F (10, 5778) = 96.36, p = < 0.001].

Before analyzing both linear regression models, scatterplots for each predictor to outcome relationship were created to confirm the linearity of each relationship. A line of best fit was applied to each scatterplot to assist with the visualization of linearity. All predictor and outcome associations were approximately linear for both linear regression models.

Residual plots were also developed for both linear regression models to check regression assumptions (see Figures 7 and 8). The linearity of each multiple regression model can be confirmed by the residual versus fitted plot. The line of best fit is approximately horizontal. Both regressions are also approximately normal as indicated by the approximately linear normal

Q-Q plots. The equality of variance assumption has also been met by both regression models since the residuals versus fitted plots do not fan out in a triangular fashion. The last assumption to check, independence, is not met by our current regression models. Responses from our various timepoints were submitted by the same set of participants at each timepoint. In order to meet the assumption of independence in future studies, regression models should be created for each data collection phase separately.

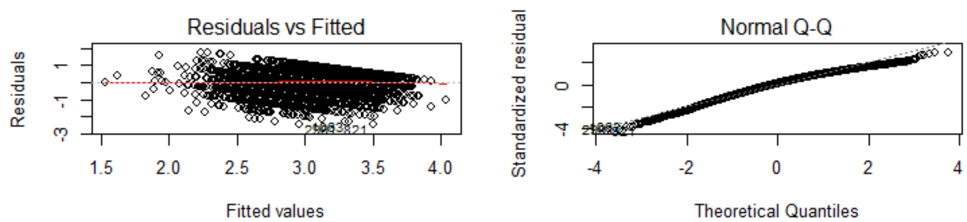


Figure 7 *Residual Plots: HSGPA Regression Model*

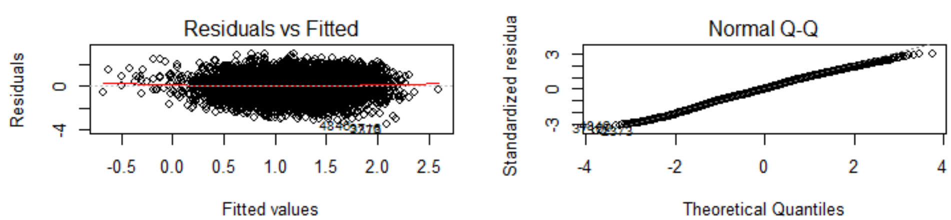


Figure 8 *Residual Plots: Mathematics Achievement Regression Model*

Each of the reduced regression models share the following predictors: BY math utility, BY science efficacy, school engagement, F1 math effort, F1 science utility, F1 science effort, and math incremental beliefs. Since these independent variables were determined to be significant predictors of multiple academic performance indicators, these common motivational constructs may be more likely to predict the presence of other academic performance indicators not measured in this thesis than other independent variables included in the full model.

The problem with this theory, however, is that most of the predictors that are present in all three regression models were assessed again at different timepoints, and all timepoints are not represented in the common reduced regression variables. For example, mathematics utility value was assessed in the BY and F1 student questionnaire; however, BY math utility value appeared in all three reduced regressions while F1 math utility only appeared in one reduced model.

Although BY math utility value may be an effective predictor for various academic performance indicators, generalizing this assumption to math utility value as a concept would be misinformed based on these results.

Tables 14-16 present coefficients for each of the three regression models used to analyze motivational constructs' ability to predict academic achievement. The first regression assessed relationships between predictors and HSGPA (see Table 14). F1 science utility value had the strongest positive association with predicted HSGPA (t = 7.336, p < 0.001). For a one standard deviation increase in F1 science utility, the predicted HSGPA increased by 0.941 points assuming all other variables remain constant. Although most predictors in this model were correlated with positive gains in HSGPA, some predictors were related to decreased HSGPA. Holding all other variables constant, as a 9th grader's perception that mathematics is important outside of school increased, the predicted HSGPA decreased by 0.103 points. F1 science efficacy (t = -3.859, p < 0.001) and mathematics incremental beliefs (t = -3.267, p = 0.001) were also associated in decreases in HSGPA when other predictors are assumed to be constant.

Table 14

Standardized Regressions Between Motivational Characteristics and High School GPA

Variable	*B*	*SE*	*t*(5778)	*p*
Constant	3.130	0.039	81.203	< 0.001
BY Math utility	-0.103	0.008	-11.721	< 0.001
BY Math efficacy	0.085	0.010	8.654	< 0.001
BY Science efficacy	0.067	0.009	7.264	< 0.001
School engagement	0.139	0.010	14.497	< 0.001
F1 Math efficacy	0.026	0.010	2.651	0.008
F1 Math effort	0.137	0.011	12.117	< 0.001
F1 Science utility	0.941	0.128	7.336	< 0.001
F1 Science efficacy	-0.038	0.010	-3.859	< 0.001
F1 Science effort	0.071	0.012	6.162	< 0.001
Math incremental belief	-0.041	0.013	-3.267	0.001

Note. Coefficients are standardized in this regression model, with µ = 0 and s = 1.

Table 15 displays results for the mathematics achievement regression model. Just as was seen in the HSGPA regression, F1 science utility value had the strongest positive association with predicted mathematics achievement (t = 6.882, p < 0.001). F1 mathematics utility also had a positive association with predicted algebraic reasoning assuming all other variables remain constant (t = 5.360, p < 0.001). BY utility values in science (t = -2.617, p = 0.009) and mathematics (t = -6.083, p < 0.001), however, were associated with decreases in mathematics achievement scores. The only other predictor related to decreases in predicted mathematics assessment scores was F1 science effort (t = -3.569, p < 0.001) when all other variables are held constant.

Table 15

Standardized Regressions Between Motivational Characteristics and Math Assessment Scores

Variable	*B*	*SE*	*t*(5778)	*p*
Constant	0.878	0.065	13.529	< 0.001
BY Math utility	-0.096	0.016	-6.083	< 0.001
BY Science utility	-0.044	0.017	-2.617	0.009
BY Science efficacy	0.245	0.016	15.550	< 0.001
School engagement	0.106	0.016	6.595	< 0.001
F1 Math utility	0.087	0.016	5.360	< 0.001
F1 Math effort	0.161	0.018	9.107	< 0.001
F1 Science utility	1.653	0.240	6.882	< 0.001
F1 Science efficacy	0.092	0.016	5.689	< 0.001
F1 Science effort	-0.069	0.019	-3.569	< 0.001
Math incremental belief	0.071	0.021	3.335	< 0.001

Note. Coefficients are standardized in this regression model, with $\mu = 0$ and $s = 1$.

The coefficients for the logistic regression model of college enrollment are presented in Table 16. Keeping consistent with results from the other two regression models in this thesis, F1 science utility value had the strongest positive association with predicted college enrollment when all other variables are assumed to remain constant (t = 3.070, p = 0.002). Only two predictors in the model were correlated with decreased predicted college enrollment—BY math utility value (t = -5.945, p < 0.001) and math incremental beliefs (t = -2.848, p = 0.004).

Table 16

Standardized Regressions Between Motivational Characteristics and College Enrollment

Variable	*B*	*SE*	*z*-value	*p*
Constant	2.406	0.200	12.041	< 0.001
BY Math utility	-0.274	0.046	-5.945	< 0.001
BY Math efficacy	0.160	0.048	3.318	< 0.001
BY Science efficacy	0.141	0.046	3.075	0.002
School engagement	0.248	0.045	5.475	< 0.001
F1 Math efficacy	0.101	0.048	2.120	0.034
F1 Math effort	0.295	0.049	5.986	< 0.001
F1 Science utility	1.882	0.613	3.070	0.002
F1 Science effort	0.141	0.047	2.973	0.003
Math incremental belief	-0.184	0.065	-2.848	0.004

Note. Coefficients are standardized in this regression model, with $\mu = 0$ and $s = 1$.

In all three regression models, school engagement was positively associated with increases in predicted HSGPA ($t = 14.497$, $p < 0.001$), mathematics achievement scores ($t = 6.595$, $p < 0.001$), and college enrollment ($t = 5.475$, $p < 0.001$) when all other variables are assumed to be held constant. This was also the case with the following predictors: *BY science efficacy* (HSGPA: $t = 7.264$, $p < 0.001$; math achievement: $t = 15.550$, $p < 0.001$; college enrollment: $t = 3.075$, $p = 0.002$), *math effort* (HSGPA: $t = 12.117$, $p < 0.001$; math achievement: $t = 9.107$, $p < 0.001$; college enrollment: $t = 5.986$, $p < 0.001$), and *F1 science utility value* (HSGPA: $t = 7.336$, $p < 0.001$; math achievement: $t = 6.882$, $p < 0.001$; college enrollment: $t = 3.070$, $p = 0.002$). More academic performance indicators not tested in this thesis may be positively related to these predictors as well, but further research would need to be conducted to support this hypothesis.

DISCUSSION AND CONCLUSION

This thesis hypothesized that students with greater perceptions of math and science utility value (1a), self-efficacy (1b), and school engagement (1c) scores will have greater academic performance (in this study, this includes higher GPAs, higher math achievement assessment scores, and increased likelihood of attending college). Although Hypothesis 1c is supported by regression analyses, there are mixed findings related to Hypotheses 1a and 1b. In the analysis of utility value, there are several findings to note. Base year utility value predictors in math and science had negative coefficient values in all three regression models. This indicates that increases in BY utility value is negatively related to academic performance, which contradicts Hypothesis 1a. However, this pattern is not seen with first follow-up utility value predictors. On the contrary, increases in F1 science utility value were associated with the largest gains in all three academic performance indicators. F1 math utility value was only significant in the mathematics achievement regression; however, its coefficient value was positive. In sum, the relationships between utility value and academic performance (1) are domain specific and (2) are contingent on when utility value is evaluated in a sample.

The mixed support surrounding the nature of utility value is supported by previous expectancy-value theory research. Whereas expectancies have a stronger relationship with performance, value components have stronger relationships with choice and persistence (Eccles, 2009; Wigfield, Tonks, & Klauda, 2016). A study conducted by Guo et al. (2018) suggests that high school trajectories of value components in multiple domains may shape each other. For example, changes in a student's perception of the importance of science may also influence their perceptions about the importance of mathematics. If students perceive science as being more important as they continue through high school, it is feasible that perceptions about the important of mathematics could shift as well. Course taking variations may be a possible explanation for these changes. For example, some high schools may require their students to take biology before chemistry/physics because chemistry/physics general requires more mathematical expertise. As science and math concepts become more intertwined throughout a high school course taking pattern, it is possible that perceptions of science and mathematics importance become more intertwined. This hypothesis was supported by our Pearson correlation analysis. The relationship between BY math and science utility value was r = .20; however, the relationship between math and science utility value grew stronger in the first follow-up phase (r = .45).

Hypothesis 1b also had mixed support through regression analyses. All BY and F1 efficacy variables included in each regression model had a positive relationship with academic performance indicators except for one predictor—F1 science self-efficacy. F1 science self-efficacy in the HSGPA linear regression model had a negative coefficient value, which indicates that increased student perception of their ability in science was related to lower total high school GPA. In the mathematics achievement regression model, however, F1 science self-efficacy had a positive coefficient value. The reasoning behind this finding may lie in Bandura's definition of student self-efficacy. He asserts that self-efficacy is a situation-specific construct (Bandura, 1977). According to Bandura (1977, p. 200), "A number of contextual factors, including the social, situational, and temporal circumstances under which events occur, enter into such [efficacy] appraisals." This statement solidifies that variation in school domains (situational) and year in school (temporal) may produce differences in efficacy expectations. Varied coursetaking patterns may be another situational factor that impacts science self-efficacy's associated with GPA.

Hypothesis 2 theorized that students with incremental views of ability are more likely to attend college. This hypothesis is not supported by the results. Biserial correlation results conclude that there is an almost non-existent relationship between holding an incremental view of intelligence and increased chances of college enrollment. Incremental belief predictors in some models would present a weak, negative regression coefficient where other models would result in a weak, positive regression coefficient. The null findings about the incremental theory of intelligence's ability to predict college enrollment may be explained by Dweck and Leggett (1988). As discussed, when an individual believes that intelligence is fixed but they perceive themselves as being intelligent, then they are more likely to produce mastery-oriented behaviors even though they possess an entity theory of intelligence. It may be the case that the participants of HSLS:09 that strongly believed that mathematics intelligence is fixed also perceive themselves as being highly intelligent, but data on perceived math intelligence in college was not collected as part of this longitudinal study.

Domain specificity of academic achievement predictors was a topic of interest in this thesis as well. Hypothesis 3 theorized that utility value (3a) and self-efficacy (3b) are domain specific in line with previous research. Hypothesis 3a was supported by paired t-test results; however, there were mixed findings about the domain specificity of self-efficacy. Based on the paired t-test results, BY self-efficacy was determined to be domain specific, but F1 self-efficacy was not. Hypothesis 4 was also related to domain specificity. It theorized that effort value would be the least domain specific predictor assessed in this thesis, and this hypothesis was supported by paired t-test results.

Buehl and Alexander (2016) discusses how teachers and students hold fundamental beliefs about the nature of each academic domain. These beliefs, known as epistemic beliefs, can shape perceptions about what is necessary to perform proficiently in those domains. If epistemic beliefs are prone to change, then it follows that student self-efficacy changes could occur. The likelihood of epistemic change lies within instructional practices. Muis and Duffy (2013) discovered that when teachers emphasize critical thinking, connections to prior knowledge, and evaluation of multiple ways to solve problems in their classrooms, epistemic beliefs shifted within the course of a semester. Classrooms that did not stress the importance of these instructional practices did not see shifts in epistemology. As a student progresses through high school, critical thinking and connections to past knowledge are necessary to help them understand course material. It can be theorized that increased utilization of critical thinking and problem-solving skills prompted epistemic change in HSLS:09 high school students, therefore transforming the nature of their efficacy beliefs over the course of their high school education.

Stability of academic performance predictors was also evaluated. It was hypothesized via Hypothesis 5 that utility value would be the most stable predictor. This hypothesis is not supported by the results. Science self-efficacy is the only construct classified as stable over time based on paired t-tests conducted. Previous research has uncovered findings that do not support this utility value hypothesis. Choiunard and Roy (2008) examined changes in high school students' mathematics utility value and found that it decreased over time. Petersen and Shibley Hyde (2015) discovered the same trajectory in their analysis of middle school mathematics utility value. Students' decreased utility value over time seems to be the hypothesized developmental trajectory of the construct according to many researchers (Fredricks & Eccles, 2002; Jacobs et al., 2002; Watt, 2004), but the significance of its decline is inconsistent in previous research. The inconsistency may be related to the domain specificity of utility value. Utility value in different domains may present different developmental trajectories.

Findings about the domain specificity and stability of each construct help us further understand the theoretical nature of each independent variable. In turn, these findings give guidance as to how these constructs should be measured. Utility value, self-efficacy, and implicit theories of intelligence were domain specific constructs in our results. Therefore, it is not appropriate to utilize generalized assessments of these constructs. Measurements that do not differentiate these constructs by their domain would be ignoring the theoretical foundations that comprise these independent variables.

Science self-efficacy was the only independent variable to be significantly stable over time. An individual's results on a science self-efficacy measurement are likely to hold steady within a small margin for at least a couple of years. However, the same assumption does not apply for the other constructs analyzed in this thesis. This idea is very important when it comes to utilizing these terms in predictive statistical models such as regression analyses. For example, an individual's assessment of utility value in 9th grade may vary significantly after a couple of years have passed. Therefore, constructs that are not considered stable over time require more frequent assessment to ensure that the evaluation of these constructs is accurate.

There are some limitations in test development and measurement that could affect results. Firstly, data about each predictor assessed in this thesis was collected using questionnaires of five or less questions. Although each of these measurements reported acceptable Cronbach's alpha values, Sijtsma (2009) states that Cronbach's alpha is dependent on the number of items in a measurement. NCES did present a brief description of their test development process; however, it is unclear how many items they started with for each motivational scale. In order to maximize construct validity during the test development process, it is imperative that researchers explicitly define their variables; however, these definitions are not included in NCES documentation. A construct definition and its theoretical framework should be discussed in-detail in order to determine which questions (and how many questions) will adequately capture each dimension of the definition.

The only reliability statistic reported by NCES for motivational scales was a weighted Cronbach's alpha value. Sijstma (2009) argues that Cronbach's alpha is not related to the internal structure of a measurement, which is defined by Bandalos (2018) as the relations among test items that mirror those expected from a theory. Rather, it provides a lower bound for test reliability of a single test administration. In order to verify evidence of internal structure, a researcher should utilize item and subscale intercorrelations, internal consistency statistics, and factor analytic procedures.

Data collected via self-report scale are prone to response biases. Bandalos (2018) identifies common biases that can come into effect when using self-report scales for noncognitive factors. Firstly, response distortion "refers to a *systematic* tendency to respond to a range of items on some basis rather than the intended content." (Bandalos, 2018, p. 104). Random responding can be one type of response distortion, and this can pose threats to the average scores for constructs measured. Socially desirable responding is also a common form of response distortion. Individuals that respond in a socially desirable manner either agree or disagree with items to portray a more positive picture of themselves. Likewise, individuals can also agree or disagree with items to portray a more negative picture of themselves. If enough participants produce socially desirable results over accurate results, this may result in an inaccurate reflection of the sample assessed and may cause researchers to make claims that are not based in reality. There are methods to manage response distortion that should be implemented to increase the possibility that item responses are more accurate.

The only exclusion criterion present in the HSLS:09 is that students must attend study-eligible schools in the United States. This means that the sample contains individuals from a variety of backgrounds, and it is possible that questionnaire items may have not been interpreted similarly across participants. To be certain that questionnaires measure the same construct across all respondents, data analysts often conduct measurement invariance testing between base year and first follow-up variables. There is no evidence that NCES conducted this testing as part of their data analysis. With such a large, diverse sample size, it is recommended to conduct measurement invariance testing before comparing composite scores to ensure that you are comparing similarly perceived constructs.

Changes in statistical analyses may provide stronger results. Although this thesis was able to uncover interesting findings about the predictive nature of motivational constructs in relation to academic performance, including interaction terms into the model may strengthen regressions and explain more of the variability present in each model. According to Montgomery, Peck, and Vining (2012, p. 69), "an interaction implies that the effect [on y] produced by changing one variable (x_1) depends on the level of the other variable (x_2)." Including interaction effects for these co-existing terms may further (1) explain the relationships between predictors and (2) better explain how the academic performance outcomes are affected by a combination of multiple motivational characteristics. Moderators and mediators were also not included in this thesis's statistical models. Both moderators and mediators are types of third variables implemented into a model with the purpose of providing a more in-depth understanding of a causal relationship between an independent variable and a dependent variable (Wu & Zumbo, 2008). Some examples of possible third variables are sample demographic characteristics, school accountability grade, and participants' past course grades.

Only one predictor, science self-efficacy, was determined to be significantly stable over time. Because many of our predictors are prone to change, utilizing latent growth curve trajectories to analyze relationships between predictors and academic performance outcomes may better capture relationships between predictors and outcomes. A latent growth curve trajectory can track whether *changes* in motivation characteristics predict performance more than a motivational construct measured at a single time point. This model has been applied to recent educational literature; Petersen and Hyde (2017) utilized this methodology in their study to examine how changes in math motivation across middle school predicts math performance in high school. Structural equation models might also be useful to adjust the results for the measurement error that is likely to be present in all construct composite scores. In order to increase the accuracy of regression coefficients, multilevel regression models or adjustments for clustering/nesting of students within schools could be incorporated into modeling. These alterations would make the standard errors of regression coefficients more accurate.

Due to the lack of exclusion criteria in the HSLS:09 sample and the similar representativeness between the complete HSLS:09 sample and this document's final analytic sample, the findings presented in this thesis can be generalized to almost any high school student in the United States. The use of random sampling techniques increases the chance that HSLS:09 is representative of the U.S. high school and college student population; however, the participants completed HSLS:09 questionnaires on a volunteer basis. This data collection style is prone to some bias innately, since "they tend to consist of individuals who are particularly invested or interested in the research study's topic." (Hanasono, 2018). Participants' reasons for taking part in the HSLS:09 were not assessed; therefore, the claim made by Hanasono (2018) cannot be definitively applied to the HSLS:09 sample unless further data is collected.

This thesis analyzed the domain specificity and stability over time for each motivational construct, adding to the limited body of research about these characteristics. Furthermore, this thesis evaluated relationships between motivational constructs and secondary/post-secondary academic performance. As discussed earlier, constructs that demonstrate adequate relationships with academic performance indicators can be utilized in academic interventions with the goal of increasing high school and college academic achievement. From this thesis, school engagement may be an appropriate construct to apply to this context. The construct itself was measured in domain-general way, so engagement interventions could theoretically be applied in the same fashion to any high school or college course. School engagement was also positively related to all three academic performance indicators assessed in this thesis.

Constructs that demonstrate increased construct stability could be employed in a college admission setting to determine which students are likely to exhibit higher rates of college retention. Science self-efficacy was the only predictor in this thesis to demonstrate construct stability over two years; therefore, this would be the recommended construct to utilize for college admissions purposes. It is important to note, however, that general measures of self-efficacy may not be appropriate in a college admissions context. Based on our results, self-efficacy was domain specific and mathematics self-efficacy was not stable over time.

Regardless of your desired application for these findings, this thesis along with many other prior studies demonstrate that the sole use of cognitive ability predictors in academic performance research excludes important variation that explains increases in academic achievement.

REFERENCES

Bandalos, D. L. (2018). *Methodology in the social sciences. Measurement theory and applications for the social sciences*. Guilford Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W H Freeman/Times Books/ Henry Holt & Co.

Bashkov, B. M., & Finney, S. J. (2013). Applying longitudinal mean and covariance structures (LMACS) analysis to assess construct stability over two time points: An example using psychological entitlement. *Measurement and Evaluation in Counseling and Development, 46*(4), 289–314. https://doi.org/10.1177/0748175613497038

Belfield, C. R., & Bailey, T. (2011). The Benefits of Attending Community College: A Review of the Evidence. *Community College Review, 39*(1), 46-68. https://doi.org/10.1177/0091552110395575

Blackwell, L. S., Trzesniewski, K. H., and Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Development*, *78*, 246–263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Bong, M. (2001). Between- and Within-Domain Relations of Academic Motivation Among Middle and High School Students: Self-Efficacy, Task-Value, and Achievement Goals. *Journal of Educational Psychology*, *93*(1), 23-34. https://doi.org/10.I037//0022-0663.93.1.23

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology, 100*(2), 431–449. https://doi.org/10.1037/a0038047

Buehl, M. M., & Alexander, P. A. (2016). Beliefs About Learning in Academic Domains. In Wentzel, K. R., & Miele, D. B. (Ed.), *Handbook of Motivation at School* (pp. 479-501). New York: Routledge. https://doi.org/10.4324/9781315773384

Chamarro-Premuzic, T., Harlaar, N., Greven, C., & Plomin, R. (2010). More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins. *Intelligence, 38*, 385-392. https://doi.org/10.1016/j.intell.2010.05.002

Chase, P., Hilliard, L., Geldhof, G. J., Warren, D., & Lerner, R. (2014). Academic Achievement in the High School Years: The Changing Role of School Engagement. *Journal of Youth and Adolescence, 43*, 884-896. https://doi.org/10.1007/s10964-013-0085-4

Chouinard, R., & Roy, N. (2008). Changes in high-school students' competence beliefs, utility value and achievement goals in mathematics. *British Journal of Educational Psychology, 78*, 31-50. https://doi.org/10.1348/000709907X197993

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Costa, A., & Faria, L. (2018). Implicit Theories of Intelligence and Academic Achievement: A Meta-Analytic Review. *Frontiers in Psychology, 9*(829). https://doi.org/10.3389/fpsyg.2018.00829

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics, 29*, 1-28. https://doi.org/10.1016/j.jhealeco.2009.10.003

De Castella, K., & Byrne, D. (2015). My intelligence may be more malleable than yours: The revised implicit theories of intelligence (self-theory) scale is a better predictor of achievement, motivation, and student disengagement. *European Journal of Psychology of Education, 30*(3), 245–267. https://doi.org/10.1007/s10212-015-0244-y

Dotterer, A., & Lowe, K. (2011). Classroom Context, School Engagement, and Academic Achievement in Early Adolescence. *Journal of Youth and Adolescence, 40*, 1649–1660. https://doi.org/10.1007/s10964-011-9647-5

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273. https://doi.org/10.1037/0033-295X.95.2.256

Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*(2), 78–89. https://doi.org/10.1080/00461520902832368

Eccles (Parsons), J. S., Adler, T. F., Futterman, R., Goff , S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco, CA: W. H. Freeman.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*(4), 519–533. https://doi.org/10.1037/0012-1649.38.4.519

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. https://doi.org/10.3102/00346543074001059

Furnham, A., & Cheng, H. (2013). Factors influencing adult earnings: Findings from a nationally representative sample. *The Journal of Socio-Economics, 44*, 120–125. https://doi.org/10.1016/j.socec.2013.02.008

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology, 52*, 1-14. https://doi.org/10.1016/j.cedpsych.2017.10.003

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Gitterman, D. P., Moulton, J. G., Bono-Lunn, D., & Chrisco, L. (2015). Can "Some College" Help Reduce Future Earnings Inequality? *Peabody Journal of Education, 90*(5), 636-658. https://doi.org/10.1080/0161956X.2015.1087774

Gunderson, E.A., Hamdan, N., Sorhagen, N.S., & D'Esterre, A.P. (2017). Who Needs Innate Ability to Succeed in Math and Literacy? Academic Domain specific Theories of Intelligence About Peers Versus Adults. *Developmental Psychology, 53*(6), 1188-1205. http://dx.doi.org/10.1037/dev0000282

Guo, J., Wang, M.-T., Ketonen, E. E., Eccles, J. S., & Salmela-Aro, K. (2018). Joint trajectories of task value in multiple subject domains: From both variable- and pattern-centered perspectives. *Contemporary Educational Psychology, 55*, 139–154. https://doi.org/10.1016/j.cedpsych.2018.10.004

Hanasono, L. K. (2018). External Validity. In M. Allen (Ed.) *The SAGE Encyclopedia of Communication Research Methods* (pp. 481-483). SAGE Publications, Inc. https://dx.doi.org/10.4135/9781483381411

Hout, M. (2012). Social and Economic Returns to College Education in the United States. *Annual Review of Sociology, 38,* 379-400. https://doi.org/10.1146/annurev.soc.012809.102 503

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*(2), 509–527. https://doi.org/10.1111/1467-8624.00421

Jiang, Y., Rosenzweig, E.Q., & Gaspard, H. (2018). An expectancy-value-cost approach in predicting adolescent students' academic motivation and achievement. *Contemporary Educational Psychology, 54*, 139-152. https://doi.org/10.1016/j.cedpsych.2018.06.005

Kiely K.M. (2014). Cognitive Function. In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_426

Kuncel, N. R., & Hezlett, S. A. (2016). Fact and fiction in cognitive ability testing for admissions and hiring decisions. In G. J. Boyle, J. G. O'Gorman, & G. J. Fogarty (Eds.), *Sage benchmarks in psychology. Work and organisational psychology: Research methodology; Assessment and selection; Organisational change and development; Human resource and performance management; Emerging trends: Innovation/globalisation/technology* (p. 123–134). Sage Publications, Inc.

Lee, J-S. (2014). The Relationship Between Student Engagement and Academic Performance: Is It a Myth or Reality? *The Journal of Educational Research, 107*, 177-185. https://doi.org/10.1080/00220671.2013.807491

Martin, J. H., Montgomery, R. L., & Saphian, D. (2006). Personality, achievement test scores, and high school percentile as predictors of academic performance across four years of coursework. *Journal of Research in Personality, 40*(4), 424–431. https://doi.org/10.1016/j.jrp.2005.02.001

Montgomery, D.C., Peck, E.A., & Vining, G.G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Wiley.

Mouratidis, A., Michou, A., & Vassiou, A. (2017). Adolescents' autonomous functioning and implicit theories of ability as predictors of their school achievement and week-to-week study regulation and well-being. *Contemporary Educational Psychology, 48*, 56-66. http://dx.doi.org/10.1016/j.cedpsych.2016.09.001

Muis, K. R., & Duffy, M. C. (2013). Epistemic climate and epistemic change: Instruction designed to change students' beliefs and learning strategies and improve achievement. *Journal of Educational Psychology, 105*(1), 213–225. https://doi.org/10.1037/a0029690

National Center for Education Statistics. (n.d.). High School Longitudinal Study of 2009: Basic Facts. https://nces.ed.gov/surveys/hsls09/basic_facts.asp

National Center for Education Statistics. (2011). *High School Longitudinal Study of 2009 (HSLS:09) Base-Year Data File Documentation* [code book]. https://nces.ed.gov/surveys/hsls09/usermanuals.asp

National Center for Education Statistics (2011, April 7). *How is Grade Point Average Calculated?* https://nces.ed.gov/nationsreportcard/hsts/howgpa.aspx

National Center for Education Statistics. (2013). *High School Longitudinal Study of 2009 (HSLS:09) Base-Year to First Follow-Up Data File Documentation* [code book]. https://nces.ed.gov/surveys/hsls09/usermanuals.asp

National Center for Education Statistics. (2015). *High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript Data File Documentation* [code book]. https://nces.ed.gov/surveys/hsls09/usermanuals.asp

National Center for Education Statistics. (2018). *High School Longitudinal Study of 2009 (HSLS:09) Base-Year to Second Follow-Up Data File Documentation* [code book]. https://nces.ed.gov/surveys/hsls09/usermanuals.asp

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*(1), 116–130. https://doi.org/10.1037/0022-3514.93.1.116

Olivier, E., Archambault, I., De Clercq, M., & Galand, B. (2019). Student Self-Efficacy, Classroom Engagement, and Academic Achievement: Comparing Three Theoretical Frameworks. *Journal of Youth and Adolescence, 48*, 326-340. https://doi.org/10.1007/s10964-018-0952-0

Perez, T., Dai, T., Kaplan, A., Cromley, J. C., Brooks, W. D., White, A. C., Mara, K. R., & Balsai, M. J. (2019). Interrelations among expectancies, task values, and perceived costs in undergraduate biology achievement. *Learning and Individual Differences*, *72*, 26-38. https://doi.org/10.1016/j.lindif.2019.04.001

Petersen, J., & Shibley Hyde, J. (2015). Trajectories of self-perceived math ability, utility value and interest across middle school as predictors of high school math performance. *Educational Psychology, 37*(4), 43-456. https://doi.org/10.1080/01443410.2015.1076765

Phan, H.P. (2009). Relations between goals, self-efficacy, critical thinking and deep processing strategies: a path analysis, *Educational Psychology, 29*(7), 777-799. https://doi-org.ezproxy.lib.purdue.edu/10.1080/01443410903289423

Pinxten, M., Marsh, H.W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology, 84*, 152-174. https://doi.org/10.1111/bjep.12028

Priess-Groben, H., & Shibley Hide, J. (2017). Implicit Theories, Expectancies, and Values Predict Mathematics Motivation and Behavior across High School and College. *Journal of Youth and Adolescence, 46*, 1318-1332. https://doi.org/10.1007/s10964-016-0579-y

Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology, 36*(4), 257–267. https://doi.org/10.1016/j.cedpsych.2011.05.002

Richardson, M., Abraham, C., Bond, R. (2012). Psychological correlates of university students academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353-387. https://doi.org/10.1037/a0026838

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis. *Psychological Bulletin, 130*(2), 261-288. https://doi.org/10.1037/0033-2909.130.2.261

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives of Psychological Science, 2*(4), 313-345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Romero, C., Master, A., Paunesku, D., Dweck, C. S., & Gross, J. J. (2014). Academic and emotional functioning in middle school: The role of implicit theories. *Emotion*, *14*(2), 227–234. https://doi.org/10.1037/a0035490

RStudio Team. (2015). *RStudio: Integrated Development Environment for R* [Computer software]. Boston, MA. http://www.rstudio.com/

Schneider, M. & Preckel, F. (2017). Variables Associated With Achievement in Higher Education: A Systematic Review of Meta-Analyses. *Psychological Bulletin, 143*(6), 565-600. https://doi.org/10.1037/bul0000098

Shively, R. L., & Ryan, C. S. (2013). Longitudinal changes in college math students' implicit theories of intelligence. *Social Psychology of Education: An International Journal, 16*(2), 241–256. https://doi.org/10.1007/s11218-012-9208-0

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist, 50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924

Stajkovic, A., Bandura, A., Locke, E., Lee, D., & Sergent, K. (2018). Test of three conceptual models of influence of the big five personality traits and self-efficacy on academic performance: A meta-analytic path-analysis. *Personality and Individual Differences, 120,* 238-245. https://doi.org/10.1016/j.paid.2017.08.014

Stefansson, K. K., Gestsdottir, S., Geldhof, G. J., Skulason, S., & Lerner, R. M. (2016). A bifactor model of school engagement: Assessing general and specific aspects of behavioral, emotional and cognitive engagement among adolescents. International *Journal of Behavioral Development, 40*(5), 471–480. https://doi.org/10.1177/0165025415604056

Trautwein, U., Marsh, H., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the Multiplicative Term in Modern Expectancy–Value Theory: A Latent Interaction Modeling Study. *Journal of Educational Psychology, 104*(3), 763-777. https://doi.org/10.1037/a0027470 van Rooij, E.C.M., Jansen, E., & van de Grift, W. (2017). Secondary school students' engagement profiles and their relationship with academic adjustment and achievement in university. *Learning and Individual Differences, 54*, 9-19. http://dx.doi.org/10.1016/j.lindif.2017.01.004

Villalón, R., Mateos, M., & Cuevas, I. (2015). High school boys' and girls' writing conceptions and writing self-efficacy beliefs: what is their role in writing performance? *Educational Psychology, 35*(6), 653-674. https://doi.org/10.1080/01443410.2013.836157

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through llth-grade Australian students. *Child Development, 75*(5), 1556–1574. https://doi.org/10.1111/j.1467-8624.2004.00757.x

Wigfield, A., & Eccles, J. (2000). Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology, 25*, 68-81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., Tonks, S., & Klauda, S. L. (2016). Expectancy-Value Theory. In Wentzel, K. R., & Miele, D. B. (Ed.), *Handbook of Motivation at School* (pp. 55-75). New York: Routledge. https://doi.org/10.4324/9781315773384

Wolfe, R. N., & Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and Psychological Measurement, 55*(2), 177–185. https://doi.org/10.1177/0013164495055002002

Wu, A. D., & Zumbo, B. D. (2008). Understanding and Using Mediators and Moderators. *Social Indicators Research, 87*, 367–392. https://doi.org/10.1007/s11205-007-914

APPENDIX

Table A1

Mathematics Utility Value Items: Student Questionnaires

Data Collection Phase	Variable Name	Questionnaire Items
Base Year (BY)	X1MTHUTI	How much do you agree or disagree with the following statements about the usefulness of your [fall 2009] math course? What students learn in this course…
		1) is useful for everyday life.
		2) will be useful for college.
		3) will be useful for a future career.
First Follow-Up (F1)	X2MTHUTI	How much do you agree or disagree with the following statements about math?
		1) Math is useful for everyday life.
		2) Math is useful for college.
		3) Math is useful for a future career.

Note. A Likert-scale format was used to collect data. Participants could respond with "strongly disagree", "disagree", "agree", or "strongly agree" for each item.,

Table A2

Science Utility Value Items: Student Questionnaires

Data Collection Phase	Variable Name	Questionnaire Items
Base Year (BY)	X1SCIUTI	How much do you agree or disagree with the following statements about the usefulness of your [fall 2009] science course? What students learn in this course…
		1) is useful for everyday life.
		2) will be useful for college.
		3) will be useful for a future career.
First Follow-Up (F1)	X2SCIUTI	How much do you agree or disagree with the following statements about science?
		1) Science is useful for everyday life.
		2) Science is useful for college.
		3) Science is useful for a future career.

Note. A Likert-scale format was used to collect data. Participants could respond with "strongly disagree", "disagree", "agree", or "strongly agree" for each item.

Table A3

Mathematics Self-Efficacy Items: Student Questionnaires

Data Collection Phase	Variable Name	Questionnaire Items
Base Year (BY)	X1MTHEFF	1) You are confident that you can do an excellent job on tests in this [fall 2009 math] course. 2) You are certain that you can understand the most difficult material presented in the textbook used in this course. 3) You are certain that you can master the skills being taught in this course. 4) You are confident that you can do an excellent job on assignments in this course.
First Follow-Up (F1)	X2MTHEFF	How much do you agree or disagree with the following statements about [math course title]/math]? 1) You are confident that you can do an excellent job on math tests 2) You are certain that you can understand the most difficult material presented in math textbooks. 3) You are certain that you can master math skills. 4) You are confident that you can do an excellent job on math assignments.

Note. A Likert-scale format was used to collect data. Participants could respond with "strongly disagree", "disagree", "agree", or "strongly agree" for each item.

Table A4

Science Self-Efficacy Items: Student Questionnaires

Data Collection Phase	Variable Name	Questionnaire Items
Base Year (BY)	X1SCIEFF	1) You are confident that you can do an excellent job on tests in this [fall 2009 science] course. 2) You are certain that you can understand the most difficult material presented in the textbook used in this course. 3) You are certain that you can master the skills being taught in this course. 4) You are confident that you can do an excellent job on assignments in this course.
First Follow-Up (F1)	X2SCIEFF	How much do you agree or disagree with the following statements about [science course title]/science]? 1) You are confident that you can do an excellent job on science tests. 2) You are certain that you can understand the most difficult material presented in science textbooks. 3) You are certain that you can master science skills. 4) You are confident that you can do an excellent job on science assignments.

Note. A Likert-scale format was used to collect data. Participants could respond with "strongly disagree", "disagree", "agree", or "strongly agree" for each item.

Table A5

School Engagement Items: Student Questionnaire

Data Collection Phase	Variable Name	Questionnaire Items
Base Year (BY)	X1SCHOOLENG	How often do you…
		1) go to class without your homework done?
		2) go to class without pencil or paper?
		3) go to class without books?
		4) go to class late?

Note. A Likert-scale format was used to collect data. Participants could respond with "never", "rarely", "sometimes", or "often" for each item.

Table A6

Effort Expenditure Items: Student Questionnaire (First Follow-Up)

School Subject	Variable Name	Questionnaire Items
Math	X2MEFFORT	How often [do/did] you do these things in [math course title]?
		1) You [pay/paid] attention to the teacher.
		2) You [turn/turned] in your assignments and projects on time.
		3) When an assignment [is/was] very difficult, you [stop/stopped] trying.
		4) You [do/did] as little work as possible; you just [want/wanted] to get by.
Science	X2SEFFORT	How often [do/did] you do these things in [science course]?
		1) You [pay/paid] attention to the teacher.
		2) You [turn/turned] in your assignments and projects on time.
		3) When an assignment [is/was] very difficult, you [stop/stopped] trying.
		4) You [do/did] as little work as possible; you just [want/wanted] to get by.

Note. A Likert-scale format was used to collect data. Participants could respond with "never", "less than half of the time", "half of the time", "more than half of the time", or "always" for each item.

Table A7

Implicit Theories of Intelligence Items: Student Questionnaire (Second Follow-Up

School Subject	Belief Type	Variable Name	Questionnaire Items
			How much do you agree or disagree with the following statements?
Math	Entity	S4MBORN	You have to be born with the ability to be good at math.
	Incremental	S4MLEARN	Most people can learn to be good at math.
Science	Entity	S4SBORN	You have to be born with the ability to be good at science.
	Incremental	S4MLEARN	Most people can learn to be good at science.

Note. A Likert-scale format was used to collect data. Participants could respond with "strongly disagree", "disagree", "agree", or "strongly agree" for each item.