

# Choosing a College STEM Major

## The Roles of Motivation, High School STEM Coursetaking, NAEP Mathematics Achievement, and Social Networks

### AIR - NAEP Working paper 2021-02

---

Jizhi Zhang

George Bohrnstedt

Xiaying Zheng

Yifan Bai

Darrick Yee

Markus Broer

**JUNE 2021**

The research contained in this working paper was commissioned by the National Center for Education Statistics (NCES). It was conducted by the American Institutes for Research (AIR) in the framework of the Education Statistics Services Institute Network (ESSIN) Task Order 14: Assessment Division Support (Contract No. ED-IES-12-D-0002/0004) which supports NCES with expert advice and technical assistance on issues related to the National Assessment of Educational Progress (NAEP). AIR is responsible for any error that this report may contain. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.



Advancing Evidence.  
Improving Lives.

*This page was intentionally left blank.*

# Choosing a College STEM Major

## The Roles of Motivation, High School STEM Coursetaking, NAEP Mathematics Achievement, and Social Networks **AIR - NAEP Working paper 2021-02**

---

JUNE 2021

**Authors:**

Jizhi Zhang, George Bohrnstedt, Xiaying Zheng, Yifan Bai, Darrick Yee, and Markus Broer



**American Institutes for Research®**  
1400 Crystal Drive, 10th Floor  
Arlington, VA 22202-3289  
202.403.5000 | [AIR.ORG](https://www.air.org)

Notice of Trademark: "American Institutes for Research" and "AIR" are registered trademarks. All other brand, product, or company names are trademarks or registered trademarks of their respective owners.

Copyright © 2021 American Institutes for Research. All rights reserved.

## AIR

Established in 1946, with headquarters in Arlington, VA, the American Institutes for Research (AIR) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of health, education, and workforce productivity. For more information, visit [www.air.org](http://www.air.org).

## NCES

The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education in the U.S. and other nations. NCES is located within the U.S. Department of Education and the Institute of Education Sciences. NCES fulfills a Congressional mandate to collect, collate, analyze, and report complete statistics on the condition of American education; conduct and publish reports; and review and report on education activities internationally.

## ESSIN

The Education Statistics Services Institute Network (ESSIN) is a network of companies that provide the National Center for Education Statistics (NCES) with expert advice and technical assistance, for example in areas such as statistical methodology; research, analysis and reporting; and Survey development. This AIR-NAEP working paper is based on research conducted under the Research, Analysis and Psychometric Support sub-component of ESSIN Task Order 14 for which AIR is the prime contractor.

William Tirre, a Program Director in the NCES Assessment Division, oversees the Research, Analysis and Psychometric Support sub-component of ESSIN Task Order 14.

### Suggested citation:

Zhang, J., Bohrnstedt, G., Zheng, X., Bai, Y., Yee, D., and Broer, M. (2021). *Choosing a College STEM Major: The Roles of Motivation, High School STEM Coursetaking, NAEP Mathematics Achievement, and Social Networks*. [AIR-NAEP Working Paper #2021-02]. Washington, DC: American Institutes for Research.

### For inquiries, contact:

**Jizhi Zhang**, Principal Researcher

Email: [jizhizhang@air.org](mailto:jizhizhang@air.org)

**Markus Broer**, Project Director for Research under ESSIN Task 14

Email: [mbroer@air.org](mailto:mbroer@air.org)

**Mary Ann Fox**, AIR Vice President and Project Director of ESSIN Task 14

Email: [mafox@air.org](mailto:mafox@air.org)

## Executive summary

Filling the STEM pipeline from secondary schools to postsecondary institutions is essential for the nations' competitiveness in the 21st century. It is therefore crucial to motivate more high school students to consider entering one of the STEM fields and to prepare themselves by taking advanced coursework in STEM that will prime them for choosing a STEM major in college, and eventually, a STEM career. High school students' pathway to a college STEM major can be seen as related to a series of academic choices (e.g., middle school STEM related activities, high school STEM coursetaking, and outside of school activities) and achievements that begin in early schooling years and continue to develop in secondary school (Eccles, 1994; Wang, 2013; Wang & Degol, 2013). This current study builds off the results of two other studies conducted under the ESSIN Task 14 contract with NCES—Mathematics Motivation and its Relationship with Mathematics Performance: Evidence from the NAEP-HSLS Overlap Sample and Examining STEM Course-taking in High School in the Prediction of Grade 12 NAEP Mathematics Scores and develops a comprehensive conceptual framework to describe how high school STEM coursetaking, STEM GPA, and motivational beliefs on science and mathematics are related to students' decisions about whether to choose a STEM major at 4-year college after taking into consideration student, family, and school background factors. The conceptual framework focuses on the direct relationships between five factors and choosing a STEM major in college: mathematics motivation, science motivation, high school STEM coursetaking, STEM achievement, and social networks.

The results of simple comparisons of mathematics and science motivation between students in STEM- and non-STEM majors indicated that STEM-major students have a higher level of mathematics and science motivation in all four measured constructs (mathematics identity, mathematics self-efficacy, science identity, and science self-efficacy) compared to non-STEM major students.

The SEM model results, which take into consideration of factors for other high school experiences, further identified the significant relationships between STEM motivation and having a STEM major in college. The findings suggest that science identity had the strongest association with students' choice of a STEM major among all other motivation variables, STEM coursetaking, and achievement variables. This is followed by mathematics identity and students' educational expectations. Students' high school STEM coursetaking and their corresponding achievement are also found to be significantly associated with the probability of students choosing a STEM major at college. The SEM results suggest that taking more credits in AP science, engineering, and overall STEM related courses along with at least 1 Physics credit is critical for choosing a STEM major in college. Students' prior STEM achievement, including NAEP mathematics scores and GPA in STEM courses, were statistically significant indicators of students' deciding to choose a STEM major in college. Students' demographic background provides social context to understand why an individual student might have chosen a STEM major in college. Female students were less likely to choose a STEM major in college compared to male students. Asian students had a higher likelihood of choosing a STEM major than White students. Regarding the economic status of a students' peers at school, the results showed that students who enrolled at schools with more than 75 percent of free or reduced-price lunch

eligible students would have a lower probability of choosing a STEM major compared to students from schools with 25 to 50 percent of students eligible for free or reduced-price lunch.

To summarize, perhaps most importantly the current study shows the central role that motivation plays in who decides to major in a STEM field in a 4-year college. Of note is the role that identities play. Thinking and believing in oneself as a scientist and as a person who is good at mathematics were shown to be key in the choice of a STEM major. Although the research literature suggested that self-efficacy is the key motivational construct for understanding who chooses a college major in a STEM subject, in the presence of measures of science and mathematics identity, self-efficacy showed no direct effect on choosing a STEM major. While this result may seem counterintuitive, perhaps it should not be. To feel efficacious about doing science or mathematics requires that one not only has the ability to solve problems in those areas but has the confidence to do so and a track record to back up that confidence. An identity in science or mathematics requires that one feels efficacious about solving problems in one or both of these areas. What distinguishes an identity is that the self—who we see ourselves as being—is comprised of our identities. According to McCall and Simmons (1978) and Stryker (1968) the self can be thought of as a hierarchy of role identities. The more prominent (important) the identity, the more core it is for the self, and the more likely we are to activate it in any given situation. It is precisely because role identities define who we are that they have such strong motivating power. It is for this reason that identities are more important in predicting role-related outcomes than self-efficacy. Self-efficacy is necessary by buttressing and supporting our role identities.

# Contents

Executive summary.....	iii
Background.....	1
Introduction.....	1
Conceptual Model.....	4
Hypothesized conceptual model .....	12
Methods .....	12
Data Source and Sample .....	12
Variables .....	14
Analyses.....	16
Results.....	19
Descriptive statistics .....	19
SEM Results.....	24
Discussion.....	29
References.....	35
Appendix A: Nested Imputations of Projected HSLs NAEP Scores .....	A-1
Appendix B.....	B-1
Appendix C.....	C-1

## List of Tables

Table 1. Measurement models for mathematics motivation and science motivation.....	18
Table 2. Student demographics, school background, and student math and STEM achievement for the final analytic sample .....	20
Table 3. Percentages of students having a STEM major at college by sex and race/ethnicity.....	21
Table 4. Means for mathematics and science motivation latent variables.....	23
Table 5. Unstandardized and standardized structural coefficients associated with entrance to 4-year college STEM major.....	28
Table A-1. Means of operational NAEP plausible values and imputed plausible values for the overlap sample .....	A-2
Table A-2. Standard deviations of operational NAEP plausible values and imputed plausible values for the overlap sample .....	A-2
Table A-3. Correlations of subscales and the composite scale using operational NAEP plausible values for overlap sample.....	A-3
Table A-4. Correlations of subscales and the composite scale using nested-imputation NAEP plausible values for overlap sample .....	A-3
Table B-1. Correlations between mathematics motivation, science motivation, STEM coursetaking, background information, and STEM major.....	B-2
Table C-1. Coefficient descriptions .....	C-2

## List of Figures

Figure 1. Conceptual framework .....	4
Figure A-1. Examples of trace-plots for means and standard deviations of one variable over iterations for MICE convergence checks.....	A-2



# Background

## Introduction

While the coronavirus pandemic will undoubtedly affect the labor market for some time to come, when the market returns, it seems likely that there will be a high demand for jobs in the field of technology, not unlike what had been true prior to the pandemic. A principal reason supporting this assumption is that Science, Technology, Engineering and Mathematics (STEM) is seen as a critical sector of the nation's economy (National Academy of Engineering & National Research Council, 2014). Prior to the pandemic, the market for science, technology, engineering, and mathematics (STEM) workers had been growing, although it has slowed somewhat in recent years. According to the Bureau of Labor Statistics (BLS), in 2014 there were about 6.1 million Americans working in STEM occupation and the projection before the pandemic was for just over 8 million workers in this sector by 2028. By far the largest growth is projected to be in the area of computer and information technology, a field which already is large in terms of current employment, and over half of the new jobs in STEM are projected to be computer or information technology related.<sup>1</sup> For example, increased use of mobile devices and the addition of software in everything from home appliances to medical devices will create demand for application software developers, with employment in this occupation projected to grow by 26 percent over the decade. Similarly, “the need for robust online security will also rise as more connected devices enter homes and workplaces. This increased need for cybersecurity will drive demand for information security analysts, employment of which is projected to grow by 31.6 percent” (Dubina, K. S., Morisi, T. L., Rieley, M., & Wagoner, A. B., 2019, p.5). So, despite recent economic challenges, there is evidence demonstrating that technology workers will continue to be in high demand as the economy recovers. However, the United States faces a major challenge in attracting, preparing, employing, and retaining a qualified workforce for jobs in the field of Science, Technology, Engineering and Mathematics (STEM).

Filling the STEM pipeline from secondary schools to postsecondary institutions is essential for the nations' competitiveness in the 21<sup>st</sup> century. It is therefore crucial to motivate more high school students to consider entering one of the STEM fields and to prepare themselves by taking advanced coursework in STEM that will prime them for choosing a STEM major in college, and eventually, a STEM career.

Several research studies have examined how postsecondary experiences impact student's persistence and attainment in STEM fields. However, few research studies have focused on how high school experiences are associated with students' entrance to a college STEM major—a critical first step to entering a STEM career (Wang, 2013) given about 80 percent of STEM college students make the decision whether to choose a STEM major in college while still in high school (Maltese & Tai, 2011; Tai et al., 2006). A deeper understanding of this first step, including the roles that motivation, peers and parents, high school STEM coursetaking, and STEM GPA play in choosing a STEM major in college, is key to building STEM support

---

<sup>1</sup> <https://www.bls.gov/emp/tables/stem-employment.htm>.

programs in high schools that would encourage and support high school students' aspirations to choose a STEM major in college and eventually a STEM occupation.

This pathway to a college STEM major can be seen as related to a series of academic choices (e.g., middle school STEM related activities, high school STEM coursetaking, and outside of school activities) and achievements that begin in early schooling years and continue to develop in secondary school (Eccles, 1994; Wang, 2013; Wang & Degol, 2013). Achievement-related behaviors such as educational and career choice are most directly associated with individuals' motivational beliefs related to STEM including interest, self-efficacy, and identity. These domain-specific motivational beliefs are shaped by an individual's social environment, such as family background, cultural norms, abilities, social networks, and their affective reactions to previous experiences as they move through adolescence towards adulthood (Eccles, 1994; Eccles et al., 1998). In other words, one's social environment (e.g., parents, peers, and teachers), individual characteristics, and experiences associated with STEM-related activities shape the development of students' identities, self-efficacy, interests, and long-term life goals, which in turn, influences educational and career choices into STEM and non-STEM fields (Eccles, 2005). Therefore, when examining who chooses STEM majors in college, it is important to understand the role that STEM motivation, high school STEM coursetaking and achievement, and students' social networks plays while also taking into account students' demographic and social backgrounds.

This report builds off the results of two other studies conducted under the ESSIN Task 14 contract with NCES. The first (Mathematics Motivation and its Relationship with Mathematics Performance: Evidence from the NAEP-HSLS Overlap Sample) examined the role of motivational factors including mathematics interest, self-efficacy, and identity along with educational aspirations (measured at grades 9 and again at grade 11), performance on a grade 9 algebra test (included in the HSLS:09 dataset), a self-report of the highest mathematics course taken by the end of students' junior year, and student and school socio-demographic data in the prediction of performance on the grade 12 NAEP mathematics assessment. Using a sample of students who not only participated in the HSLS:09 survey, but also grade 12 NAEP, a longitudinal structural equation model results demonstrated the important role that having a mathematical identity at grade 11 has on grade 12 mathematics achievement net of the other factors.

The second report (Examining STEM Course-taking in High School in the Prediction of Grade 12 NAEP Mathematics Scores) examined the roles that STEM coursetaking and STEM GPA play in the prediction of grade 12 NAEP mathematics achievement, also using the overlap sample of HSLS:09 participants who had their coursetaking and GPA data included in the High School Transcript Study and took the grade 12 NAEP assessment in 2013. The results showed that regarding coursetaking, the most important predictors were the number of Advanced Placement (AP) or International Baccalaureate (IB) credits earned in mathematics and science; precalculus, calculus or AP/IP calculus as the highest mathematics courses taken; and having earned at least one credit in either chemistry or physics while taking into account race/ethnicity, SES and gender.

Using a national representative sample of the High School Longitudinal Study (HSLS:09) data and 2015 NAEP grade 12 mathematics assessment data, this study develops a comprehensive

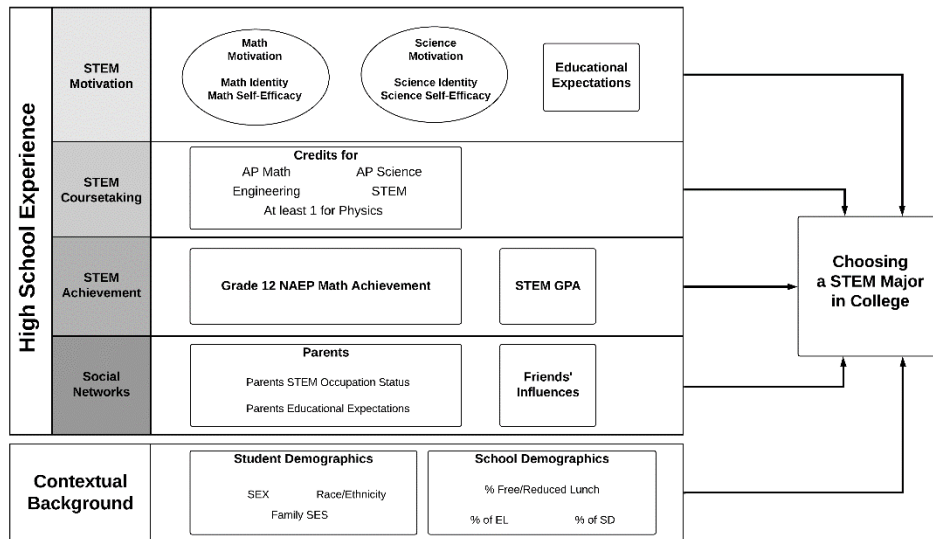
conceptual framework to describe how high school STEM coursetaking, STEM GPA, and motivational beliefs on science and mathematics are related to 4-year college students' decisions about whether to choose a STEM major after taking into consideration student, family, and school background factors. The conceptual framework focuses on the direct relationships between five factors and choosing a STEM major in college: mathematics motivation, science motivation, high school STEM coursetaking, STEM achievement, and social networks (see Figure 1).

As part of a study examining the role of mathematics motivation in predicting grade 12 mathematics achievement, Zhang et al. (2018) found that motivational beliefs (i.e., mathematics identity, mathematics self-efficacy, and mathematics interests), are positively related among themselves (both cross-sectionally and longitudinally) when examined at grades 9 and 11, but that only grade 11 motivational beliefs were directly related to grade 12 mathematics performance when both the grade 9 and grade 11 motivational measures were included in a structural equation model. Given this finding, the current study focuses on examining grade 11 STEM motivational measures along with high school STEM coursetaking, STEM achievement, and students' social networks in the prediction of choosing a STEM major in college.

More specifically, this study addresses the following four research questions:

1. What is the relationship between high school students' STEM motivation including educational expectations and their choice of a STEM major in college taking into account family and school background factors, high school STEM coursetaking, STEM achievement, and students' social networks?
2. What is the relationship between high school students' STEM coursetaking, and their choice of a STEM major in college taking into account family and school background factors, STEM motivation, social networks, and STEM achievement?
3. What is the relationship between high school students' STEM achievement, measured by students' STEM GPA and NAEP grade 12 mathematics scores, and their choice of a STEM major in college taking into account family and school background factors, STEM motivation, social networks, and STEM achievement?
4. What is the relationship between high school students' social networks and their choice of a STEM major in college taking into account family and school background factors, STEM motivation, high school STEM coursetaking, and STEM achievement?

Figure 1. Conceptual framework



NOTE: EL refers to English learner. SD refers to students with disabilities.

## Conceptual Model

### *Motivation<sup>2</sup>*

Motivation can be thought of as the drive or energy expended in pursuit of a goal (e.g., choosing a STEM major). It can either be extrinsic (based on actual or perceived rewards or punishments) or intrinsic (based on internalized beliefs and values about the importance of achieving the goal) (Deci, 1975). More generally, motivation has been defined as “the process whereby goal-directed activity is instigated and sustained” (Pintrich & Schunk, 2002, p. 5). Because one cannot observe this internal latent process, a theoretical framework is necessary for measuring motivational beliefs and can be “conceptualized in different ways, with different theories focusing on different psychological processes” (Lameva & Chonteva, 2013, p. 4).

Much of contemporary literature on achievement motivation is grounded in the socio-cognitive perspective, focusing on understanding students’ cognitive processes with an emphasis on whether they choose to engage in learning tasks and how persistent they are while working on those tasks considering the interrelationships of individual’s motivational beliefs and environmental factors (Eccles & Wigfield, 2002).

Modern motivational theories such as the social cognitive career theory (SCCT) and expectancy-value theory have provided insights on how to understand students’ academic and career choices in general, and more specifically, students’ decisions about entering a STEM field (Lent et al., 2010). The SCCT is based on Bandura’s social cognitive theory (Bandura, 1977, 1986, 1997) which underlines how interrelationships of individual psychological constructs and family and social background influence individuals’ academic and career choices. Self-efficacy beliefs, expectations, and environmental factors are key constructs in SCCT. Similarly,

<sup>2</sup> Some of the literature review in this study draws on the literature review in Zhang et al (2018).

expectancy-value theory (EVT) emphasizes individuals' cognitive processes of academic and career choices, with a focus on the role of individuals' expectancies for success beliefs (competence-related beliefs which are similar to Bandura's self-efficacy beliefs) and values related to their academic engagements (Eccles et al., 1983). Competence-related beliefs, value beliefs, and interests are the core motivational constructs in EVT. According to the expectancy-value theory and numerous empirical studies, individuals' expectancy beliefs (e.g., self-efficacy and expectancies about success) lead to deeper cognitive engagement on a task and higher levels of performance, whereas value beliefs (e.g., interests and subjective values) influence tasks/activities that individuals choose to pursue (Wigfield & Eccles, 2000). For example, if students hold a high level of self-efficacy and value towards STEM fields, they are more likely to participate in STEM-related tasks or to enroll in more advanced STEM-related courses during their schooling years. Furthermore, the individual tends to display higher level of perseverance when challenges are encountered.

Overall, motivational theories explain individuals' academic performance and decision-making through the interplay of psychological constructs and social contexts. Specifically, a large body of research has identified self-efficacy as a particularly influential and parsimonious motivational construct to explain students' academic performance and the choice of a STEM major in college (Jenson et al., 2011; MacPhee et al., 2013; Rittmayer & Beier, 2008).

More recently, researchers have demonstrated that students' identity belief is another important motivational factor that plays a significant role in influencing individual's academic behaviors, achievement, and career choice (Eccles, 2009; Stets et al., 2017; Zhang, et al., 2018). Individuals' behaviors and decisions (e.g., choosing a STEM major at college) are influenced by personal and social identities as well as influencing the expression of personal and social identities. In addition, individuals' identities and their self-efficacy beliefs have a reciprocal relationship, where ones' identity can influence their self-efficacy beliefs and their self-efficacy beliefs can influence their identity.

Therefore, the current study integrates self-efficacy and role identity into the conceptual framework for understanding high school students' decision to choose a STEM major in college. The current study hypothesizes that both identity and self-efficacy are positively associated with the likelihood of high school students entering a STEM field in college.

## **Identity**

In educational settings, identity has long been considered an important part of students' academic success and persistence (Marsh, 1990, 1993; Marsh et al., 1988). More recently, researchers have studied the impact of social identities and role identities (e.g., mathematics identity, science identity) on academic and career choice, including how these identities relate to minority and female students' academic choices in STEM-related fields (Chemers et al., 2011; Eccles, 2009; Stets et al., 2017; Syed et al., 2011).

A social identity is defined as a person's belief that he or she belongs to a social category or group (Abrams & Hogg, 1988; Stets & Burke, 2000). For example, individuals classify their social identities through their gender or racial/ethnic group. Eccles (2009) extended the expectancy-value model by giving attention to the role of social identity in influencing

individual's academic behaviors and performance. She focuses on the role of personal social/collective identity and contends that individuals' identities can influence and be influenced by their efficacy beliefs or their expectation for success. This reciprocal relationship was demonstrated by Zhang and her colleagues (2018). Further, Eccles argues that individuals' behaviors and decisions (e.g., choosing advanced STEM courses at high schools) are influenced by personal and collective identities as well as influencing the expression of personal and collective identities. That is, the relationship is a reciprocal one. For example, gender as a collective identity has been applied to explain why women are underrepresented at the STEM fields. Being a female as a collective identity comes with a stereotyped status, values, and role-expectations, which impacts women's academic behaviors and career choices.

While social identities are important in shaping individuals' beliefs as well as their academic and career choices, Stets and her colleagues argue that role identities also are significant (2017) and demonstrated that an individual's science identity positively impacts the likelihood of entering a science occupation. In role identity theory, the core of an identity is the categorization of the self as an occupant of a role (e.g., a scientist or a mathematician), and the incorporation, into the self, of the meanings and expectations associated with that role and its performance (Burke & Tully, 1977; Thoits, 1986; Stets & Burke, 2000). From the perspective of the role identity theory, whether one choose a STEM major in college is dependent upon the degree to which one embraces a science or math identity. Measures of a science role identity, for example, would ask whether one sees themselves as a science person and whether others see them as a science person. If the answer is in the affirmative, it is hypothesized that the person would then seek to demonstrate this by taking increasingly more advanced STEM courses in high school and would choose a STEM major in college. Zhang et al. (2018) using data from the HSLs:09-NAEP overlap sample found that, among high school students, having a mathematics identity was significantly associated with NAEP mathematics performance after taking into consideration their mathematics self-efficacy, mathematics interests, educational expectations, and demographic backgrounds. This study takes a symbolic interactionist perspective, in which identities are a function of the meanings that people attach to the roles that they play—student, father, scientist, and so on (Stryker & Burke, 2000). The self is seen as composed of a hierarchy of identities where the more prominent or salient the identity, the higher it is in the hierarchy (McCall & Simmons, 1978; Stryker, 1968). Following Stone (1962), identities are established when significant others use the same words to describe someone as that person uses for themselves. Thus, to be *identified*, claims made for oneself must be legitimated and supported by significant others. For example, a student's science identity is based not only on their own perceived capabilities and accomplishments in science, but also through the support of this perception by others.

How parents, teachers, and friends view someone in relation to a subject (i.e., math or science) has an impact on that person's perceptions of their own competence in that subject (Bleeker & Jacobs, 2004). For example, Wenger (1998) found that one's perception of his or her mathematics identity was impacted by others in the community, which then influenced one's participation within that community. Furthermore, they found others can identify skills and abilities in individuals, leading to new identity claims not previously discovered and demonstrating the reciprocal nature of identity development. In a similar vein, individual student's science identity is established through personal claims that are confirmed by significant others in one's life.

Not surprisingly, research shows that students' role identity plays a significant role in their academic achievement and career choice. Eccles and Barber (1999) found that underrepresented minorities with a strong role identity had greater persistence to degree completion compared to those with strong social identities (e.g., based on race/ethnicity or gender). Carlone and Johnson (2007) developed a model of science identity to understand pathways to STEM careers of 15 successful minority women. They conceptualized an individual's science identity in terms of three components: competence; performance; and recognition. More specifically, they theorized that individuals with higher levels of self-efficacy beliefs regarding their ability to master STEM related subjects and content will be able to better carryout STEM-related tasks and jobs, reaffirming a science identity through personal experience and feedback. Their results highlighted that the recognition of one's science identity plays a significant role in moving women of color through the STEM pipeline from postsecondary STEM majors to STEM occupations after graduation.

Chemers et al. (2011) sampled 665 undergraduate and graduate students and found that having a science identity significantly predicted students' commitment to science careers. Stets and her colleagues (2017) utilized national panel data that followed underrepresented college students in STEM fields from 2005 to 2013 to investigate how an individual's science identity in the educational system influences the choice to enter a science occupation. They found that having a science identity was positively associated with students' likelihood of entering a science occupation after taking into consideration of other control factors including science self-efficacy, STEM GPA, science activities, and demographic background.

Not only do mathematics and science identities grow out of one's previous accomplishments in mathematics and science as well as from recognition from others for those accomplishments, they are also related to feeling efficacious about doing mathematics and science, as both an antecedent and consequence of having a mathematics or science identity. These in turn leads to additional mathematics and science pursuits, such as an increased number of challenging mathematics and science courses, and the eventual decision to pursue a college degree in a STEM field.

## **Self-Efficacy**

Bandura (1994) defined self-efficacy as: people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives. Self-efficacy beliefs determine how people feel, think, motivate themselves, and behave. Such beliefs produce these diverse effects through four major processes. They include cognitive, motivational, affective, and selection processes (p. 71).

Zimmerman (2000) summarized that (1) Self-efficacy measures focus on performance capabilities rather than personal qualities, such as physical or psychological capabilities; (2) Self-efficacy beliefs are multidimensional and differ on the basis of the domain of functioning (e.g., self-efficacy beliefs about performing on a mathematics test may differ from beliefs about performance on a reading test); (3) Self-efficacy beliefs depend on a mastery criterion of performance rather than on normative or other criteria (e.g., students evaluate how well they are at solving an arithmetic problem, not how well they expect to do on the problem

compared with other students); and (4) Self-efficacy beliefs specifically refer to future performance and are assessed before students perform the relevant activities.

In Eccles's expectancy-value theory, self-efficacy beliefs are referred to as competence-related beliefs and are of two types: ability beliefs and expectancy beliefs (Eccles et al., 1983; Wigfield & Eccles, 2000). *Ability beliefs* are defined as individuals' perceptions of their *current* competence, particularly in a specific domain such as mathematics or science. These beliefs reflect evaluations not only of their own ability, but how their ability compares to others. *Expectancy beliefs* refer to how one thinks he or she will do on *future* tasks, either in the immediate or longer-term future. Expectancy-value theorists hypothesize that expectancies for future performance are influenced by ability beliefs, the perceived difficulty of different tasks, individual goals, and previous experiences. For example, individuals' beliefs about their mathematics ability comes from many years of experiences with mathematics and reflect their own evaluation of their current skills in mathematics. Individuals' expectancies refer to how they think they will do in the future (e.g., in an upcoming mathematics course), and they base their expectancies of success primarily on their beliefs about their ability in math. As important as these two constructs are theoretically, empirical research has shown that they cannot be easily be differentiated (Eccles & Wigfield, 1995; Eccles et al., 1993). Eccles and Wigfield (2002) concluded that, "[a]pparently, even though these constructs can be theoretically distinguished from each other, in real-world achievement situations they are highly related and empirically indistinguishable" (p. 119).

The formulation of self-efficacy through expectancy-value, with its emphasis on ability and expectancy beliefs, supports Bandura's self-efficacy theory, given its stress on individuals' beliefs about their capabilities to generate performances that have influence over events that impact their lives.

In modern symbolic interaction theory, self-efficacy is increased by self-verification of an identity in that it leads to a sense of control over the role-related environment (Stets & Burke, 2000). However, that sense of control also motivates behaviors that provide the opportunity to verify identities. Ervin and Stryker (2001) postulated that self-efficacy is both an antecedent as well as a consequence of role-related identity behaviors. Brenner et al. (2018) examined the reciprocal relationship between role-specific self-efficacy, identity prominence, and identity salience using four waves of data from a longitudinal study of underrepresented college students in STEM fields. Identity prominence is seen as the affective component of identity; that is, it is the subjective value assigned to an identity (McCall & Simmons, 1978) while identity salience is defined as the probability that a given identity will be enacted in a particular situation or setting (Stryker, 1968, 2004). Brenner et al. (2018) found that science self-efficacy was related to identity prominence, which in turn was related to identity salience. While the direct relationships between identity prominence and salience and role-specific self-efficacy over time were statistically significant, both relationships were weak.

Self-efficacy has been considered an essential motivational construct to understand student's academic achievement and behaviors (Zimmerman, 2000). Extensive research indicates that students with lower mathematics self-efficacy or lower perceived competence in mathematics perform less well on mathematics tests compared to students with higher mathematics self-efficacy (Eccles, 2009; Valian, 2007; Wigfield et al., 2006; Zimmerman & Kitsantas, 1997,



1999). In addition, students with higher perceived competence in mathematics are more likely to enroll in advanced mathematics courses, participate in mathematics clubs, and choose a quantitative college major (Dweck, 2008). Similarly, students' science self-efficacy beliefs positively influence students' science achievement (Britner, 2008; Britner & Pajares, 2006; Cohen & Chang, 2020). Britner (2008) studied how self-efficacy impacted high school science students' achievement in life, physical, and Earth science subject areas. Students' self-efficacy beliefs were a significant predictor of science grades across three subjects, though the positive relationship differed by gender. Cohen and Chang (2020) used Trends in International Math and Science Study 2011 data to investigate how science self-efficacy, value in learning science, science attitudes, gender, and race impact science achievement among United States middle school students. Their finding revealed that self-efficacy is positively related to students' science achievement, but the level of self-efficacy varied among gender and racial/ethnic groups.

In addition, self-efficacy has been linked to students' college major and career choices, particularly in the STEM areas (Lent et al., 1986, 1991; Pajares, 1996; Williams & George-Jackson, 2014). Researchers have reported that the mathematics self-efficacy of college students has a larger impact on their choice of math-related courses and majors than their prior mathematics achievement (Hackett, 1985; Hackett & Betz, 1989; Lent et al., 1991; Pajares & Miller, 1994, 1995). Lent et al. (1991) examined mathematics self-efficacy and its relation to career selection in science-related fields for college students. The results revealed that mathematics self-efficacy predicts interest and course selection in mathematics-related areas. Sahin et al. (2017) studied 2,246 high school graduates from a STEM-focused public school in Texas and found that high school students with higher levels of mathematic self-efficacy and science self-efficacy are 1.33 times and 1.37 times more likely to select a STEM major in college, respectively. Williams & George-Jackson (2014) studied 1,881 college students whose majors were in STEM-related fields. They found that "students' confidence and beliefs about their own abilities and skills have a much greater impact on using their scientific skills and knowledge than the extent to which they see themselves as a scientist, or to the extent that there is a gender-based interaction" (p. 115).

Overall, it can be inferred that having a strong sense of self-efficacy in mathematics and science significantly impacts students' choices of entering STEM majors in college and eventually having career success in STEM. Both students' role identity and self-efficacy have important implications for students in science- and mathematics-based majors. For the current study, student role-identity and self-efficacy will serve as important lenses of analysis to understand the extent to which both constructs impact students' decision of entering a STEM major in college.

## **Educational Expectations**

In addition to the two motivation variables noted above, adolescents' expectations about their future educational attainment are also important as they influence academic choices, decisions, behaviors, and activities as well as academic achievement (Nurmi, 2004; Zhang, et al., 2018). High school seniors' educational aspirations as part of the Wisconsin Longitudinal Study of 1957 were the strongest predictor of educational outcomes seven years later (Gasson et al., 1972; Sewell et al., 1970; Sewell & Shah, 1968). Additionally, Bozick et al. (2010) used data from the Wisconsin Longitudinal Study to show that stable educational expectations across individuals' school careers are important predictors not only of school performance but also of long-term

status attainment. Ou and Reynolds (2008) investigated the relationship between student educational expectations and educational attainment among 1,286 low-income, minority students who grew up in an urban area and found that students' expectations were one of the strongest predictors of educational attainment. Lichtenberger and George-Jackson (2013) investigated factors that influenced student's early interest in taking a STEM major in college using a random sample of the Illinois High School Graduating Class of 2003. They found that a higher level of student educational expectations increased the likelihood of having an early interest in STEM.

Researchers also suggest that students' educational expectations influence their academic-related decisions and activities, such as taking a higher-level STEM course in high school, thus shaping their academic achievement and ultimate educational attainment. Beal and Crockett (2010) conducted a longitudinal study of 317 adolescents and found that the students' educational expectations significantly predicted their likelihood of participation in academic activities (e.g., school science fair, math club).

### ***STEM course learning experiences and STEM achievement***

Empirical evidence has demonstrated that high school math and science coursetaking is influential in high school students' pursuit of a STEM major in college and eventually a career in STEM disciplines (Trusty, 2002; Federman, 2007; Tyson et al., 2007; Rask, 2010; Gottfried, 2015). Tyson et al. (2007) found that Florida students who took calculus courses in high school were more likely to obtain STEM degrees in college compared to those who did not, although this same pattern was not followed for students who took Algebra II as their most advanced math course as compared to students who only took Algebra I or Geometry. Similarly, students who completed Chemistry II or Physics II were more likely to get a STEM degree than students who did not take these courses, and the role of high school Physics courses in particular was important in influencing the future attainment of a STEM focused degree. The finding revealed that:

Students in the Physics I category obtain STEM degrees at 18.7 percent, above the 14.0 percent the rate for all baccalaureate degree recipients. The Chemistry II or Physics II group produces the largest percentage of STEM students at 39.8 percent, but this was only the case for 72 of 425 students. Physics course-taking is a primary factor in STEM attainment. Almost three quarters of baccalaureate degree recipients (1,693, 72.8 percent) took Physics I. This suggest some advantage over students who took Chemistry, even though 1,417 (61.0 percent) of students who took Chemistry completed a STEM bachelor's degree. Still, only 8.8 percent of students who took Chemistry I, but not Physics I completed a STEM bachelor's degree (p. 258).

This study emphasizes the importance and value of taking high-level coursework in both mathematics and science during high school for the completion of a STEM degree in college.

Rask (2010) investigated attrition among STEM majors by using college administrative data for the graduating classes of 2001–2009 from a Northeastern liberal arts college with approximately 5,000 graduates. The findings indicated that college students' pre-collegiate preferences in high school played an important role in STEM attrition rates. Both the number of AP credits and high school intended college major were the most consistent and important influences on the decision to complete a STEM major. The results suggest that focusing on high school preparation

(especially AP credits) and preferences for majoring in a STEM discipline in college would not only increase the number of students declaring a STEM major in college, but also play a significant role in helping students persist to graduation with a STEM major.

Gottfried (2015) used the Education Longitudinal Survey (ELS:2002) data to investigate how high school students' mathematics, science, and applied STEM coursework is associated with the likelihood of students declaring a STEM major in college. The study found that high school students who took more advanced mathematics courses (e.g., Trigonometry and Calculus), were more likely to choose a STEM major in college than students taking less advanced mathematics courses. Similarly, students who took more advanced science courses (e.g., Advanced Chemistry and Physics) had higher odds of declaring a STEM major in college. The study also found that taking applied STEM courses (e.g., Information Technology and Scientific Research and Engineering) was associated with higher odds of declaring a STEM major in college. Overall, research findings suggest that both academic and applied STEM learning in high school appear to be key factors in increasing probability of declaring a STEM college major and graduating from college with a STEM degree. Research also indicates that students' academic achievement in STEM related subjects has a positive relationship with students' likelihood of entering a STEM major in college. Radunzel et al. (2017) investigated around 91,000 four-year college students across 43 postsecondary institutions in the United States who were first-time college enrollees between the years 2005 and 2009. They found that higher levels of mathematics and science achievement in high school increased the likelihood of students declaring a STEM major in college.

### **Social networks**

Students' social networks play an important role in shaping their behaviors and decisions. Parents and peers are typically the two most important social networks for high school students. Both parents and friends have significant influence on students' school performances and college decisions (Fehrmann et al., 1987; Hossler & Stage, 1992).

Research has also shown that parental expectations are positively correlated with students' educational success (Benner & Mistry, 2007; Catsambis, 2001). In addition, parental expectations are more influential on students' college choices than teacher or peer expectations (Brasier, 2008; Ma, 2001). According to Rutchick et al. (2009), parental expectations also change students' expectations for themselves. Their longitudinal studies found that parental educational expectations measured at the beginning of the study continue to impact students' academic performance even five years later. Archer et al. (2013) found that aspirations in STEM occupations are also impacted by parents' attitudes. Sahin, Ekmekci and Waxman (2017) investigated more than 2,000 high school students from Texas and found that students whose parents had a college degree are more likely to major in STEM fields than those whose parents did not. Moakler and Kim (2014) used national freshmen survey data and found that students were more likely to choose STEM majors if their parents are in STEM occupations. In summary, students' STEM academic and career choices are directly impacted by parental expectations, attitudes, and occupations (Lee et al., 2015; Leppel et al., 2001).

Friends are another important component of students' social network—a component that plays a role in influencing students' academic and career choices. Riegle-Crumb, Farkas, and Muller

(2006) investigated how friend groups impacted high school male and female students' advanced coursetaking. They found that female students' friend groups had a strong impact on their advanced-coursetaking patterns, especially in the areas of mathematics and science, which inevitably influence their decisions to choose a STEM major in college. Raabe, Boda, and Stadtfeld (2019) investigated large-scale panel data on adolescents from Sweden and found that students adjust their preferences for their education and career to those of their friends. For example, female students tend to keep their STEM preferences when their friends also like STEM.

## **Hypothesized conceptual model**

This study's hypothesized conceptual model takes a comprehensive perspective in investigating the direct relationships between high school STEM learning experiences and the likelihood of students entering a STEM major in college, with an emphasis on motivational beliefs. Four aspects of high school experiences are included in the model: (1) STEM motivation; (2) STEM coursetaking; (3) STEM achievement; and (4) Social networks.

The current framework is embedded in the Social Cognitive Career (SCC) theory, which incorporates students' motivational beliefs, academic behaviors, and their social and contextual background into one model to help understand high school students' choice of a STEM major in college.

## **Methods**

### **Data Source and Sample**

This study uses two datasets: (a) the full HSLs:09 data from multiple waves including Base Year (2009), First Follow-up (2012), 2013 Update (2013), High School Transcript (2013–2014), and Second Follow-up (2016); and (b) the special overlap sample of approximately 3,480 students who participated in the High School Longitudinal Study of 2009 (HSLs:09) and also took the 2013 Grade 12 National Assessment of Educational Progress (NAEP) Mathematics Assessment. In addition, a two-stage multiple imputation was conducted to compute the projected NAEP mathematics achievement scores for the full HSLs sample members. First, the imputation was implemented for the full HSLs sample members' background variables selected in the current study. Therefore, the full HSLs sample members' imputed NAEP plausible values were able to be computed by conditioning on all HSLs variables. Then, a latent regression imputation model was used to generate the NAEP plausible values for the full HSLs sample members (Ogut, Bohrnstedt & Broer, 2015).

The HSLs:09 base year took place in the 2009–10 school year, with a randomly selected sample of fall term ninth graders from a sample of over 900 schools and over 25,000 students. Students took an algebra assessment and a survey online in their schools. The first follow-up of HSLs took place in the spring of 2012 when most participants were in the spring of their eleventh grade. The first follow-up (2012) survey covered topics such as students' STEM motivation, mathematics and science extracurricular participation, and school experiences. The 2013 Update collected information on students' college plans and choices, college applications, financial aid

applications and employment experiences. High School Transcripts (2013–2014) collected students high school grades, including STEM and AP courses taken. The second follow-up was conducted in 2016, when most sample students were three years beyond high school graduation. It collected students' postsecondary education and employment data, including self-reported STEM major status and degree attainment status.

The missing values of key variables included in the conceptual model were imputed for the full HSLs sample using the Multivariate Imputation by Chained Equations (MICE) procedure. Twenty sets of imputations were conducted for the current study. Appendix A describes details of the techniques used for handling missing values.

As a result of imputation, the NAEP-HSLs overlap sample provides an opportunity to examine the predictive validity of grade 12 NAEP mathematics scores for college enrollment and the choice of a STEM major. The current study investigates whether NAEP scores are associated with the probability of students' entrance to a STEM major in college, as well as the direct relationships between other HSLs variables, including motivational variables and STEM coursetaking variables, and the likelihood of students entering a STEM field in college. Grade 12 NAEP mathematics score is an overall score for mathematics, which is a composite of five subscales: (a) number properties and operations; (b) measurement; (c) geometry; (d) data analysis, statistics, and probability; and (e) algebra. NAEP scale scores are reported as "plausible values" because, by the NAEP assessment design, students are administered only a small subset of the total pool of assessment items, not representing the whole assessment domain.<sup>3</sup> Multiple imputation procedures are then used to produce a set of 20 "plausible values" (e.g., plausible scores) for each student taking the mathematics assessment. In generating the plausible values, NAEP uses a "conditioning model" that includes all the variables from the various contextual questionnaires NAEP collects along with the responses of students to the particular portion of the assessment items that they are assigned. The proper analysis of NAEP data requires that the variables included in the analysis are also included in the conditioning model. This requirement posed a problem for the analyses in this study given that many of the variables came from the HSLs questionnaires. A previous study conducted by AIR demonstrated that using the NAEP plausible values not conditioned on variables from the HSLs contextual student questionnaire produced biased results (Zhang et al., 2018). Therefore, a new set of 20 plausible values conditioning on all variables used in the current study for the full HSLs sample was produced for the 20 imputed datasets from the MICE procedure. As a result, a total of 400 datasets were created for the final analyses. The details of the imputation procedure for the NAEP plausible values can be found in Appendix A.

There were 13,433 students who participated in all waves of the study and who also had transcript data available. Because the study uses multiple waves of the HSLs survey, longitudinal weights, made available by NCES, were applied for the analyses to account for differential selection probabilities and patterns of nonresponse at each wave of the study. The current study investigates students' choice of STEM major in college, therefore the HSLs sample students who did not report enrollment in a 4-year postsecondary college between high

---

<sup>3</sup> For more information, see NAEP technical documentation on the web: <http://nces.ed.gov/nationsreportcard/tdw/analysis>.

school graduation and February 2016 were excluded from the final analytic sample. Only 4-year college enrollees were used because an exploratory analysis indicated 2-year college students differ in demographics, achievement, and motivation from 4-year college students, indicating that separate analyses are needed for 4-year and 2-year college students. Since only students who had enrolled in a 4-year college were used, the final analytic sample contained around 6,820 students.

## **Variables**

This section summarizes outcome and independent variables that were used in the current study based on the hypothesized conceptual framework.

### ***Outcome Variables***

The main outcome variable is students' choice of major at 4-year colleges. It is a dummy variable with 1 indicating a STEM major and 0 representing a non-STEM major. STEM majors can include a wide range of disciplines. For example, the National Science Foundation (NSF) identifies STEM majors as mathematics, natural sciences, engineering, computer and information sciences, psychology, economics, sociology, and political science. However, recent federal and state legislative efforts have focused on improving STEM education mainly in mathematics, natural sciences, engineering, and technology (Kuenzi et al., 2006). One such effort with a more restrictive STEM definition is called the National Science and Mathematics Access to Retain Talent (SMART) definition; it excludes the social and behavioral sciences.<sup>4</sup> More specifically, only the following fields were designated as STEM majors by SMART<sup>5</sup>: Computer Sciences; Engineering; Life Sciences; Mathematics; Physical Sciences; and Technology.

To identify who was a STEM major, if a student reported a major in a STEM field as the degree they were working towards at any time during their postsecondary enrollment between high school graduation and February 2016, they were considered a STEM major (including, in the case of double majors, one of the two majors being a STEM major). This was true whether the student was actively working towards a degree in February 2016, had completed it, or had stopped pursuing it.

### ***Independent Variables***

**STEM Motivation.** Students' psychological beliefs are hypothesized to play an important role in impacting students' decisions to enter a STEM major in college, including students' STEM motivation and their educational expectations.

The current study focused on motivation related to two specific subject areas: mathematics and science, which arguably are the most essential STEM disciplinary areas. Students' mathematics

---

<sup>4</sup> Details about SMART Grants STEM eligible majors can be found at <https://www2.ed.gov/rschstat/eval/highered/smart-grant/acg-smart-grant-report-year-third-final.pdf>.

<sup>5</sup> SMART Grants STEM eligible majors also include agriculture, natural resources, and multidisciplinary studies. However, the current study did not include these majors.

and science motivation were measured both at grade 9 and 11 in HSLs:09, with items assessing students' subject-specific identity and self-efficacy. A previous AIR NAEP research study has investigated how mathematics motivation evolved longitudinally from grade 9 to grade 11 and its impacts on students' grade 12 mathematics achievement (Zhang et al., 2018). That study found that when both grade 9 and 11 mathematics identity and self-efficacy were in the model, only the grade 11 motivational variables were significantly associated with grade 12 mathematics achievement. Therefore, the current study targeted the role of grade 11 (the most recent follow-up) students' STEM motivational beliefs on students' entrance to a STEM major at college. Given that the same wording was used for motivation items in both mathematics and science questionnaires, only mathematics identity and mathematics self-efficacy items are described in detail.

To measure mathematics identity, students were asked to report how strongly they agreed with the following statements: (1) You see yourself as a mathematics person; and (2) Others see you as a mathematics person. The items were coded as (0) Strongly disagree, (1) Disagree, (2) Agree, and (3) Strongly agree. For mathematics self-efficacy, students were asked how strongly they agreed with the following statements: (1) You are confident that you can do an excellent job on tests in this 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; and (4) You are confident that you can do an excellent job on assignments in this course. All items were coded as (0) Strongly disagree, (1) Disagree, (2) Agree, and (3) Strongly agree.

In addition, students' educational expectations were measured at grade 11 by asking "As things stand now, how far in school do you think you will get?" The responses for this question were recoded as (1) Less than Associate's degree, (2) Completion of Associate's degree to less than Bachelor's degree, (3) Completion of Bachelor's degree to less than Master's degree, (4) Completion of Master's degree to less than Ph.D./M.D./law degree/high level professional degree, and (5) Completion of Ph.D./M.D./law degree/other high level professional degree.

**STEM coursetaking.** The second part of the conceptual framework focuses on students' STEM coursetaking behaviors in high school. Four items were included to capture students' total Carnegie credits<sup>6</sup> earned in (1) AP and IB Mathematics courses, (2) AP and IB Science courses, (3) Engineering and Technology courses, and (4) STEM courses overall. Additionally, a dichotomized item was used to indicate whether students earned at least one Carnegie unit in Physics. Physics was chosen since it is generally regarded as the most difficult of the various science courses that a high school student could take, which was also supported by the data.

**STEM achievement.** In terms of students' high school STEM achievement, the analyses included the student's NAEP mathematics score at grade 12 in 2013 and the student's grade point average (GPA) in STEM courses. The NAEP grade 12 mathematics score, theoretically

---

<sup>6</sup> The Carnegie unit is a measure of the amount of time a student has studied a subject. A total of 120 hours in one subject earns the students one "unit" of high school credit (see <https://www.carnegiefoundation.org/faqs/carnegie-unit> for details).

ranging from 0 to 300, consists of 20 plausible values per student while student's GPA in STEM courses ranged from 0 to 4 points.

**Social networks.** Given the conceptual model, variables measuring students' social networks focus on their parents' and friends' influences. Parental educational expectations were derived from the first follow-up parent questionnaire in HSLs:09 asking parents how far in school they thought their child would go. The responses indicated the highest level of education parents expected the child to achieve. The categories were grouped and recoded as: (1) Less than Associate's degree, (2) Completed Associate's degree but less than Bachelor's degree, (3) Completed Bachelor's degree but less than Master's degree, (4) Completed Master's degree but less than Ph.D./M.D./law degree/high level professional degree, and (5) Completed Ph.D./M.D./law degree/high level professional degree.

Whether a parent had a STEM occupation or not was also hypothesized to influence the choice of a STEM major. Therefore, a dummy variable was generated to indicate if either parent's current or most recent job was in a STEM occupation when surveyed at the baseline or at the first follow-up. Occupations in Life and Physical Science, Engineering, Mathematics, and Information Technology are categorized as STEM occupations. This categorization of STEM occupations follows the Occupational Information Network (O\*NET) taxonomy. The O\*NET Program is the nation's primary source of occupational information, which is developed under the sponsorship of the U.S. Department of Labor/Employment and Training Administration through a grant to the North Carolina Department of Commerce.

Finally, a latent variable was created to capture the influence of a student's friends. Students were asked whether the following items were true or false for their closest friend at grade 9, including (1) gets good grades, (2) is interested in school, (3) attends classes regularly, and (4) plans to go to college.

**Student demographic background.** This study employed students background information from HSLs:09 data. The student's gender was coded as a dummy variable (i.e., 0 = male, 1 = female). Meanwhile, several dummy variables were generated to indicate the student's race/ethnicity, including White, African American, Hispanic, Asian, or Other. Finally, the analysis included the student's family socioeconomic status (SES), a composite variable calculated by the HSLs:09 using parent/guardian's education, occupation, and family income when the student was in grade 9.

**Student school context.** Variables used to measure students' school contexts were derived from the baseline HSLs:09 school questionnaire. School level poverty was measured as the percentage of students eligible for free or reduced-price lunch. The variable was recoded into three categories: (1) 0-25 percent, (2) 26-75 percent, and (3) more than 75 percent. The study also used two other variables to measure school context: (1) percentage of English learners (EL) and (2) percentage of students with disabilities who received special education services.

## Analyses

This section describes the analytical methods adopted to answer the study's research questions. The overall model used is structural equation modeling (SEM), and the model is estimated using



Version 8.3 of Mplus. The SEM model has two elements: a measurement model and a structural model. The measurement model is used to verify how well the items measure the four motivational constructs they are presumed to measure. The structural part of the model examines how well each of the independent variables in the model (including the four motivational latent variables) relate to whether students who go to 4-year colleges choose a STEM major by 2016 or not.

**The Measurement Model.** Confirmatory Factor Analysis (CFA) was used to assess the fit of the items to the four latent motivational variables (factors) that they are presumed to measure (science and mathematics identity; science and mathematics self-efficacy). The presumption is that each item loads on only the factor it is presumed to measure and none of the other three. Results from the CFA shown in Table 1 indicate an excellent fit of the items to the hypothesized 4-factor model: a comparative fit index (CFI) = 0.995, Tucker-Lewis index (TLI) = 0.994, Root mean square error of approximation (RMSEA)=0.038. The standardized factor loadings of the items ranged from 0.92 to 0.93 for mathematics identity, 0.84 to 0.94 for mathematics self-efficacy, 0.90 to 0.97 for science identity, and 0.84 to 0.94 for science self-efficacy. These relatively large factor loadings suggest that the items were good measures of the latent constructs, supporting their inclusion in the structural model. The intercorrelations among the latent variables ranged from 0.20 to 0.66 as found in Appendix B.

Table 1. Measurement models for mathematics motivation and science motivation

	Standardized factor loading ( $\lambda$ )	Standard error
Mathematics motivation		
Mathematics identity		
ID1: You see yourself as a mathematics person	0.92	0.02
ID2: Others see you as a mathematics person	0.88	0.02
Mathematics self-efficacy		
SE1: You are confident that you can do an excellent job on tests in this course	0.92	0.01
SE2: You are certain that you can understand the most difficult material presented in the textbook used in this course	0.83	0.01
SE3: You are certain that you can master the skills being taught in this course	0.91	0.01
SE4: You are confident that you can do an excellent job on assignments in this course	0.90	0.01
Science motivation		
Science identity		
ID1: You see yourself as a science person	0.91	0.02
ID2: Others see you as a science person	0.94	0.02
Science self-efficacy		
SE1: You are confident that you can do an excellent job on tests in this course	0.92	0.01
SE2: You are certain that you can understand the most difficult material presented in the textbook used in this course	0.83	0.01
SE3: You are certain that you can master the skills being taught in this course	0.91	0.01
SE4: You are confident that you can do an excellent job on assignments in this course	0.90	0.01

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

**Structural Model.** To answer the research questions, structural equation modeling was used to examine the relationships between students’ STEM motivation, STEM coursetaking behaviors, STEM achievement, social networks, students’ demographic background, school background and students’ declaring a STEM major. The final model includes both the measurement and structural components. To estimate parameters, a Weighted Least Square (WLSMV) estimator was used. For the measurement model, the variance of each latent construct was fixed to one and the first indicator was estimated freely (Muthén & Muthén, 2018).

Given the dependent variable—STEM major or not—is a dichotomous variable, a probit structural model was used where the inverse standard normal distribution of the probability of choosing a STEM major is modeled as a linear combination of the predictors:

$$P(Y_i = 1) = \Phi(\beta_0 + \beta_1 \text{Mathidentity} + \beta_2 \text{Mathselfefficacy} + \beta_3 \text{Scienceidentity} + \beta_4 \text{Scienceselfefficacy} + \beta_5 \text{EDEpectation} + \beta_6 \text{APmath} + \beta_7 \text{APscience} + \beta_8 \text{Engineering} + \beta_9 \text{STEM} + \beta_{10} \text{Physics} + \beta_{11} \text{NAEP} + \beta_{12} \text{GPAstem} + \beta_{13} \text{ParentalSTEMOCC} + \beta_{14} \text{ParentalEDEpectation} + \beta_{15} \text{Friends} + \beta_{16} \text{Female} + \beta_{17} \text{Black} + \beta_{18} \text{Hispanic} + \beta_{19} \text{Asian} + \beta_{20} \text{Other} + \beta_{21} \text{SES} + \beta_{22} \text{Schoollunch} + \beta_{23} \text{SchoolELL} + \beta_{24} \text{SchoolSpecEd} + \varepsilon_i)$$

where  $Y_i$  represents choice of STEM or non-STEM major,  $\beta_0$  is the intercept, and  $\beta_1$  to  $\beta_5$  are the path coefficients for the four motivation latent constructs (mathematics identity, mathematics self-efficacy, science identity, and science self-efficacy), and students' educational expectations, respectively.  $\beta_6$  to  $\beta_{10}$  are the structural coefficients for students' coursetaking variables, including students' credits earned in AP mathematics courses, credits earned in AP science courses, credits earned in Engineering and Technology courses, credits earned overall in STEM courses, and whether students earned at least one credit in Physics. Continuing,  $\beta_{11}$  to  $\beta_{12}$  are the corresponding structural coefficients for students' NAEP mathematics score at grade 12 in 2013 and students' GPA in STEM courses.  $\beta_{13}$  to  $\beta_{15}$  are the structural coefficients for parent STEM occupation, parental educational expectations, and friends' social influence. Next, the structural coefficients for students' background, including female, Black, Hispanic, Asian, Other race/ethnic group, and SES, are estimated as  $\beta_{16}$  to  $\beta_{21}$ . Finally, for school characteristics,  $\beta_{22}$  to  $\beta_{24}$  are the estimated coefficients for percentage of students eligible for free or reduced-price lunch at the school, percentage of EL students and percentage of students with disabilities who received Special Education services. Finally,  $\varepsilon_i$  is the disturbance term. This disturbance term is assumed to be uncorrelated with any of the independent variables in the model. Appendix C presents the detailed information regarding the path coefficients and the disturbance term of the SEM model.

## Results

### Descriptive statistics

#### *Student and school demographics*

Table 2 presents descriptive statistics for the student- and school-level demographic background variables for the 6,820 students enrolled in 4-year colleges. Among these students, 45 percent were male, and 55 percent were female. In terms of race/ethnicity, 62 percent were White, 11 percent were African American, 15 percent were Hispanic, 6 percent were Asian, and 7 percent indicated another race. With respect to school characteristics, 31 percent of students received free or reduced-price lunch, 5 percent were English learners, and 11 percent had a disability. The mean GPA for STEM courses taken in high school was 2.96 and the mean score for the dependent variable, grade 12 NAEP mathematics assessment, was 169.10.

Table 2. Student demographics, school background, and student math and STEM achievement for the final analytic sample

Demographic characteristics	Weighted percentage/mean	
	Percentage	Standard error
Student demographic characteristics		
Sex		
Male	45	1.10
Female	55	1.10
Race/ethnicity		
White	62	1.60
Black	11	1.00
Hispanic	15	1.10
Asian	6	0.60
Other	7	0.50
School background		
Mean school free/reduced lunch	31	1.00
Mean school EL students	5	0.41
Mean school students with disabilities	11	0.36
	Mean	Standard error
Student achievement		
STEM GPA	2.96	0.02
NAEP mathematics score	169.10	0.92

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLs:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016).

Table 3 compares demographic background between students who chose non-STEM majors and those in STEM majors. Among male students, 34 percent chose STEM majors and 66 percent non-STEM majors. For female students, only 18 percent chose a STEM major. That is, percentagewise, roughly half as many females chose a STEM major compared to males. In terms of differences in choosing a STEM major by race/ethnicity, Asian American students were more likely to choose a STEM major compared to students from other race/ethnic backgrounds: 42 percent of Asian students chose a STEM field while only 26 percent of White, 20 percent of African American, and 23 percent of Hispanic students chose a STEM major.

Table 3. Percentages of students having a STEM major at college by sex and race/ethnicity

Student demographic characteristics	Students who enrolled in 4-year college			
	STEM major		Non-STEM major	
	Weighted percentage		Weighted percentage	
	Percentage	Standard error	Percentage	Standard error
Sex				
Male	34	1.50	66	1.50
Female	18	1.20	82	1.20
Race/ethnicity				
White	26	1.10	74	1.10
Black	20	2.89	80	2.89
Hispanic	23	3.02	77	3.02
Asian	42	4.16	58	4.16
Other	23	3.20	77	3.20

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

### ***STEM motivation***

Table 4 compares mathematics and science motivation between students in STEM and non-STEM majors. Two types of comparisons are presented here—one using the latent scores and the other using indices based on the item raw scores. First, when comparing students’ differences in factor means for mathematics and science motivation (i.e., identity and self-efficacy), STEM major students’ scores were set to be 0 to facilitate the comparison with non-STEM major students. Setting the means to a fixed value for one of the groups is required since latent variables do not have a natural metric. As expected, the results indicate that students in non-STEM majors had significantly lower scores on all four motivation latent variables compared to STEM majors. For example, the latent mean scores for mathematics identity and mathematics self-efficacy for non-STEM major students were  $-0.64$  and  $-0.57$ , respectively, compared to 0 for STEM majors. Since the variances were set to zero for all four latent variables, the standard deviations are 1.0. As a result, the interpretation of the results is straightforward: non-STEM majors score nearly two-thirds of a standard deviation lower on the identity measure than STEM majors and nearly three-fifths of a standard deviation lower on the mathematics self-efficacy measure. Similarly, the non-STEM majors score nearly two-thirds of a standard deviation lower than the STEM majors on the latent science identity measure and just over two-fifths of a standard deviation lower on the science self-efficacy latent variable. All the differences are statistically significant at the 0.05 level.

Second, the raw scale for each item related to motivation ranged from 0 to 3. Indices for each measure were constructed by adding students’ scores on the items comprising each of the scales and dividing by the number of items measuring the construct (e.g., two for each of the identity measures). Constructing indices using the raw score metric allows one to determine whether the average for any given group is more in the “agree” or “disagree” direction. Given the scale scores range from 0 to 3, the mid-point between agreeing and disagreeing is 1.5. Any score above 1.5 is in the direction of agreement, and any score below 1.5 is in the direction of disagreement.

The right panel of Table 4 shows that the average scores of all motivation constructs for non-STEM majors were significantly lower than those for STEM majors. For example, STEM majors scored 2.06 on the mathematics identity index, which suggests that students on average agreed that they are “a mathematics person” and others see them as a mathematics person as well. By comparison, for non-STEM majors, the average score on the index was 1.44, suggesting on average students disagreed slightly with the two statements. The results for mathematics self-efficacy are very similar with means of 2.06 and 1.47, respectively. For science identity and science self-efficacy, the pattern of mean differences is the same as for mathematics identity and mathematics self-efficacy (the mean for STEM majors is larger than that for non-STEM majors); however, note that the means for science identity and science self-efficacy for the non-STEM majors are 1.78 and 1.84 – both in the *agree* direction. That is, both the STEM and the non-STEM majors see themselves as “science persons” (and so do others); STEM majors are just stronger in their beliefs.

Overall, STEM majors had stronger beliefs about their mathematics and science identities and self-efficacy than non-STEM majors.

Table 4. Means for mathematics and science motivation latent variables

	Students in 4-year college (latent)						Students in 4-year college (raw)					
	STEM major		Non-STEM major				STEM major		Non-STEM major			
	Mean	SD	Mean	SD	P value	S.E.	Mean	SD	Mean	SD	P value	S.E.
Mathematics motivation latent variables												
Grade 11 mathematics identity	0.00	0.79	-0.64	0.80	<0.001	0.04	2.06	0.80	1.44	0.89	<0.001	0.04
Grade 11 mathematics self-efficacy	0.00	0.90	-0.57	0.88	<0.001	0.05	2.06	0.76	1.47	0.82	<0.001	0.03
Science motivation latent variables												
Grade 11 science identity	0.00	0.79	-0.64	0.81	<0.001	0.04	2.18	0.63	1.78	0.68	<0.001	0.03
Grade 11 science self-efficacy	0.00	0.92	-0.43	0.90	<0.001	0.04	2.14	0.67	1.84	0.72	<0.001	0.03

NOTE: Factor means for STEM major students' mathematics and science motivation motivations (i.e., identity, self-efficacy, and interest) were set to 0 to facilitate the comparisons of factors means between STEM major students and non-STEM major students.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

## SEM Results

To answer the research questions, structural equation modeling was used to examine the relationships between students' STEM motivation, STEM coursetaking behaviors, STEM achievement, social networks, students' demographic background, school background and students' choice to pursue a STEM major.

Table 5 presents the unstandardized and standardized structural coefficients that are associated with choosing a STEM major in 4-year colleges. The fit of the data to the model was excellent with CFI = 0.98, TLI = 0.98, and RMSEA = 0.022<sup>7</sup>. To facilitate the interpretation of the results, marginal probability effects for some of the key variables for the current study including motivation and STEM coursetaking were calculated, while all other variables were held constant at their means.

**Research questions 1:** What is the relationship between high school students' STEM motivation including educational expectations and the choice of a STEM major in college taking into account family and school background factors, high school STEM coursetaking, STEM achievement, and students' social networks?

First, students' science and mathematics identity were both statistically associated with students' choice of STEM major, with corresponding standardized coefficients of 0.24 and 0.13. Statistically speaking, this means that for a one standard deviation increase in science identity, there would be an 8-percent increase in the probability of choosing a STEM major, when all other variables included in the model equal their means. Similarly, the standardized coefficient of 0.13 for mathematics identity suggests that a one standard deviation increase in mathematics identity corresponds to about a 4-percent increase in choosing a STEM major for an otherwise "average" student.

By contrast, controlling for the other variables in the model, students' mathematics self-efficacy and science efficacy had very small and non-significant associations with students' choice of a STEM major, with standardized coefficients of 0.06 and 0.00, respectively. Correspondingly, for a one standard deviation change in mathematics self-efficacy, there is a net 2-percent change in the likelihood of choosing a STEM major—a difference which is small and not significant—and no difference as a function of science efficacy.

Students' educational expectations were hypothesized to be an important aspect of intrinsic motivation and the results from the probit analysis support that hypothesis with a significant coefficient of 0.07. This result can also be interpreted as students with higher level of educational expectations have a higher chance of choosing a STEM major at a 4-year college.

**Research question 2:** What is the relationship between high school students' STEM coursetaking, and their choice of a STEM major in college taking into account family and school background factors, STEM motivation, social networks, and STEM achievement?

---

<sup>7</sup> The model fit statistics were from one of the 400 imputed datasets.



Regarding students' coursetaking behaviors in high school, results suggest that taking more credits in AP science, engineering and overall STEM related courses is critical for choosing a STEM major in college. The standardized coefficients for credits in AP science, engineering and all STEM courses taken were 0.12, 0.08, and 0.14, respectively. These coefficients correspond to net 4-percentage, 3-percentage, and 4-percentage point increases in the probability of choosing a STEM major for a one standard deviation increase in credits taken in AP science, engineering, and all STEM credits combined, respectively, when all other variables included in the model are equal to their means.

Somewhat surprisingly, the number of credits taken in AP mathematics was not strongly related to choosing a STEM major as shown by the small and non-significant standardized coefficient of 0.01. It is surprising because previous literature has shown a significant relationship between taking credits in AP mathematics and the choice of a STEM major (Gottfried, 2015). It is possible that this relationship is suppressed by other variables related to credits taken in AP mathematics in the model.

In contrast, students who took at least 1 credit in Physics in high school were more likely to choose a STEM major in college than those who did not. The corresponding standardized coefficient was 0.09, which can be transformed and interpreted as follows: compared to those who did not take any credits in Physics, students who took at least 1 credit would have a 3-percent greater likelihood of choosing a STEM major.

**Research question 3:** What is the relationship between high school students' STEM achievement, measured by students' STEM GPA and NAEP grade 12 mathematics scores, and their choice of a STEM major in college taking into account family and school background factors, STEM motivation, social networks, and STEM achievement?

The model also includes students' prior STEM achievement including NAEP mathematics scores and GPA in STEM courses. Both variables were statistically significant indicators of students' deciding to choose a STEM major in college. Their respective standardized coefficients were 0.13 for NAEP achievement and 0.15 for STEM GPA in STEM courses. In terms of probability, students with a one standard deviation increase in their NAEP score would have a 4-percent increase in the likelihood of choosing a STEM major. Similarly, one standard deviation increase in students' STEM GPA is associated with a 5-percent increase in the likelihood of choosing a STEM major.

**Research question 4:** What is the relationship between high school students' social networks and their choice of a STEM major in college taking into account family and school background factors, STEM motivation, high school STEM coursetaking, and STEM achievement.

In term of students' social networks, neither parental nor peer factors were significantly associated with students' choice of a STEM major after controlling for other variables in the model. The standardized coefficients for at least one of the parents having a STEM occupation, parental educational expectations, and friends' influence were 0.04, 0.05, and -0.02, respectively.

For students' demographic background, the study showed some significant differences in choosing a STEM major as a function of students with varying demographic characteristics. Female students, as found in prior literature, were significantly less likely to choose a STEM major than male students. Regarding race/ethnicity differences, Asian students were more likely to choose STEM major than students from other racial groups. Meanwhile, there were no significant differences between White, African American, Hispanic and Other students in the probability of choosing a STEM major after controlling for all other variables in the model. Another finding was that students' family SES as measured by HSLs:09 was not a significant predictor of students' choice of a STEM major after controlling for other factors.

Finally, at school level, the results suggested that students in schools with more than 75 percent of students eligible for free/reduced-price lunch were less likely to choose a STEM major than students who came from schools with 26 to 50 percent of their students receiving free or reduced-price lunch. However, this result should be interpreted with caution because the distribution of this variable was quite skewed with a very small number of cases for some of the categories. A school's percentage of EL students and its percentage of students with special disability were not significantly associated with students' choice of STEM major in college given both coefficients were near zero.

### ***Summary of findings***

The simple comparisons of mathematics and science motivation between students in STEM- and non-STEM majors indicated that STEM-major students have a higher level of mathematics and science motivation in all four measured constructs compared to Non-STEM major students. Non-STEM majors score nearly two thirds of a standard deviation lower on the identity measure than STEM majors and nearly three fifths of a standard deviation lower on the mathematics self-efficacy measure. Similarly, the non-STEM majors score nearly two thirds of a standard deviation lower than the STEM majors on the latent science identity measure and over two fifths of a standard deviation lower on the science self-efficacy latent.

The SEM model results, which take into consideration of factors for other high school experiences, further identified the significant relationships between STEM motivation and having a STEM major in college. In conclusion, when examining to what extent students' motivation influences students' choice of a STEM major, the findings suggest that science identity had the strongest association with students' choice of a STEM major with a standardized coefficient of 0.24. For a one standard deviation change in students' science identity, there is an 8-percent point difference in the probability of choosing STEM major when all other variables included in the model are equal to their means. This is followed by mathematics identity with a standardized coefficient of 0.13 and students' educational expectations with a standardized coefficient of 0.07. Respectively, these two variables demonstrate 4 percent and 2 percent differences in the probability of choosing STEM major.

Students' high school STEM coursetaking and their corresponding achievement are also found to be significantly associated with the probability of students choosing a STEM major at college. The SEM results suggest that taking more credits in AP science, engineering and overall STEM related courses along with at least 1 Physics credit is critical for choosing a STEM major in college. Students' prior STEM achievement, including NAEP mathematics scores and GPA in

STEM courses, were statistically significant indicators of students' deciding to choose a STEM major in college.

Students' demographic background provides social context to understand why an individual student might have chosen a STEM major in college. Female students were less likely to choose a STEM major in college compared to male students. Asian students had a higher likelihood of choosing a STEM major than White students. Regarding the economic status of a students' peers at school, the results showed that students who enrolled at schools with more than 75 percent of free or reduced-price lunch eligible students would have a lower probability of choosing a STEM major compared to students from schools with 25 to 50 percent of students eligible for free or reduced-price lunch.

Table 5. Unstandardized and standardized structural coefficients associated with entrance to 4-year college STEM major

	Unstandardized coefficient	S.E.	p	Standardized coefficient	S.E.	p
Entrance to STEM major						
Motivation						
Mathematics identity	0.16	0.04	<0.001	0.13	0.04	<0.001
Mathematics self-efficacy	0.07	0.04	0.10	0.06	0.04	0.10
Science identity	0.28	0.05	<0.001	0.24	0.04	<0.001
Science self-efficacy	0.00	0.04	0.96	0.00	0.04	0.96
Educational expectations	0.08	0.04	0.04	0.07	0.03	0.03
STEM coursetaking						
Credits AP math	0.01	0.07	0.88	0.01	0.04	0.88
Credits AP science	0.20	0.05	<0.001	0.12	0.03	<0.001
At least 1 credit Physics	0.22	0.08	<0.001	0.09	0.03	<0.001
Credits engineering	0.15	0.04	<0.001	0.08	0.02	<0.001
Credits STEM	0.08	0.02	<0.001	0.14	0.03	<0.001
STEM achievement						
NAEP mathematics score	0.01	0.00	0.01	0.13	0.05	0.01
STEM GPA	0.25	0.06	<0.001	0.15	0.03	<0.001
Social network						
Parental STEM occupation	0.13	0.08	0.10	0.04	0.02	0.10
Parental educational expectations	0.05	0.03	0.09	0.05	0.03	0.09
Friends' influences	-0.03	0.04	0.53	-0.02	0.04	0.52
Demographic background						
Female	-0.43	0.07	<0.001	-0.36	0.05	<0.001
Black	0.20	0.12	0.08	0.17	0.10	0.08
Hispanic	0.15	0.12	0.20	0.13	0.10	0.20
Asian	0.32	0.14	0.02	0.27	0.12	0.02
Other	0.07	0.11	0.54	0.06	0.09	0.53
Family SES	-0.07	0.05	0.15	-0.04	0.03	0.14
School background						
Percentage of students with free/reduced price lunch 0–25%	-0.03	0.07	0.69	-0.02	0.06	0.69
Percentage of students with free/reduced price lunch >75%	-0.54	0.25	0.03	-0.45	0.21	0.03
Percentage of EL students	0.00	0.06	0.95	0.00	0.05	0.95
Percentage of students with disabilities	0.05	0.12	0.67	0.04	0.10	0.67

NOTE: For gender, male was the reference group. For race/ethnicity, White was the reference group. For percentage of students with free/reduced lunch, category of 25 percent to 50 percent was the reference group. The standardized coefficients for dummy coded variables (Black, Hispanic, Asian/Pacific Islander, Other, Gender, and Percentage of students with free/reduced lunch) were based on STDY standardization.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

## Discussion

Using nationally representative high school student's data, this study tested out a comprehensive conceptual framework to understand how high school experiences are related to students' academic choice of entrance to a STEM major. Structural equation model results point to important aspects of high school experiences related to students' entrance into STEM fields including students' motivational beliefs, high school STEM coursetaking and STEM achievement. The limited measures of family and student influences that were available in HSL:09 and could be included in the model were not related to the choice of a STEM major net of the other variables in the analysis.

### ***Mathematics Identity and Science Identity***

Of the four motivational constructs (mathematics identity and self-efficacy; science identity and self-efficacy) measured at grade 11, both mathematics identity and science identity were related positively to students' choice of a STEM major in college. Of the two, students' science identity was the more important; indeed, it had the largest coefficient among all of the variables examined in the study. This finding highlights the significant role that academic identity plays for funneling more high school students into the postsecondary STEM pipeline. Academic identities (e.g., science identity, mathematics identity) have recently received increased attention in the literature, especially for their role in promoting under-represented students' involvement in STEM fields. In this study, the measure of identity focused on seeing oneself as "a math/science person" and having that view supported by others. That is, one not only sees oneself as being "a math/science person," but one is also identified by significant others (e.g., teachers, parents, friends) as "a math/science person." The impact of content-specific identity on students' academic behaviors, decisions, and achievement have been reported by other researchers. For example, Stets et al. (2017) found that students' science identity was the only factor predicating whether students entered a science occupation after taking other variables in the model into account—variables such as students' science self-efficacy, GPA, and other demographic background characteristics. The current study's findings are in line with those of the Stets et. al. study in that mathematics identity and science identity were also the only motivation variables that were significantly associated with the likelihood of students entering a STEM major in college when taking into consideration of other motivation variables.

This finding renders important practical implications. Role-specific identities (e.g., math identity, science identity) are important for motivating role-specific behaviors that can further reinforce that identity. Students with a science identity are likely to take more STEM related courses and participate in STEM related activities, which in turn appears to improve their STEM achievement. STEM achievement, in turn, increases the probability of choosing a STEM major which likely improves the probability of choosing a STEM occupation.

Some teaching practices have been demonstrated to be effective in enhancing students' academic identities. The Mathematical Agency Improvement Community (MAIC), a diverse network of K-12 schools in Southern California, has adopted student-centered math practices to teach underserved middle and high school students. The student-centered practices aim to create learning cultures that nurture mathematical identity and success. Practices highlighted under the MAIC program include, among others, the launching, exploring, and discussing of lesson

structure with students, anticipatory planning, students working in small groups, and students presenting their thinking.<sup>8</sup> The MAIC classrooms are different from conventional classrooms that often focus on “one right way” of working through problems (i.e., the way taught by the teacher). In MAIC classrooms, the “right way” is of less importance. Students in the MAIC classrooms are given the chance to share and develop their own mathematical thinking and problem-solving methods which allows them to discover the relevance of mathematics which in turn can lead to the development of a mathematics identity.

The current study did not find statistically significant net relationships between either science or mathematics self-efficacy and students’ likelihood of choosing a STEM major in college. This finding is in contradiction with other studies (Jenson et al., 2011; MacPhee et al., 2013; Rittmayer & Bejer, 2008; Wang, 2013) in which self-efficacy has been highlighted as the most powerful and parsimonious factor that impacts students’ decision in choosing a STEM major in college. Furthermore, students’ self-efficacy belief is the key motivation variable in the SCCT framework. However, these studies did not include identity-related variables in their analyses. Bohrstedt et al. (2020) found that students’ mathematics identity has a larger effect on students’ mathematics achievement compared to students’ self-efficacy beliefs when both self-efficacy and identity were included in the model. In addition, Bohrstedt et al. (2020) investigated the longitudinal relationships between math identity and math self-efficacy and found that motivational beliefs are positively associated with each other and that the effect of mathematics self-efficacy on mathematics achievement was indirect through mathematics identity. Stets et al. (2017) summarized the relative importance of the two motivational constructs this way: “Our research shows that among all the factors considered, the identity process is the primary mechanism through which minority students choose a science occupation. Where other theoretical processes, like goal theory, self-efficacy theory, attitude theory, have effects, they operate through the identity process” (p. 17).

### ***Educational Expectations***

In addition to students’ sense of mathematics and science identity, the current study also found that students’ educational expectations were positively related to the likelihood of choosing a STEM major. Adolescents’ expectations about their future educational attainment are important because they influence academic choices, decisions, behaviors, and activities as well as their academic achievement. Previous research has established a link between students’ educational expectations and academic achievement, having an early interest in having a STEM major in college, and the successful completion of a college degree (Bozick, Alexander, Entwisle, Dauber, & Kerr, 2010; Lichtenberger & George-Jackson, 2013). The current study extends research on the significant impact of students’ educational expectations on academic outcomes to include students’ choice of a STEM major. While student expectations are likely not a sufficient cause for students to choose a STEM major, they may be a necessary one. That is, just because a student has high educational expectations is unlikely to lead a student to choose a STEM major in college. However, it appears that high educational expectations are common among those who do choose a STEM major.

---

<sup>8</sup> More details about the MAIC practices can be found at <https://www.mathagency.org/network-research-findings>.

Educational expectations, self-efficacy and identity are all motivators, but educational expectations appear to operate differently from the other two. As Bohrnstedt et al. (2020) note:

“Educational expectations may be a *general motivator* while identity and self-efficacy are *role-related motivators*. As a result, one would predict that educational expectations will be related to taking whatever courses are perceived as necessary for the educational outcome aspired to. That is, educational expectations get translated into coursetaking since coursetaking is seen as instrumental to achieving one’s educational achievement goals (e.g., getting a master’s degree). By way of contrast, role-related identity (and self-efficacy) motivate one in a narrower way than educational expectations do...” In the current study, the motivation in “a narrower way” would be in pursuit of a STEM-related major.

Unlike students’ educational expectations, the current study did not find a significant relationship between parents’ educational expectations for their adolescent child and students’ choice of a STEM major in college. And, the study found that having a parent with a STEM occupation was also unrelated to choosing a STEM major as were the academically related behaviors of their peers. As noted in the literature review, other researchers have documented the importance of the attitudes and behaviors of one’s peers. It should be noted that the measure of peer influences in this study did not include items that asked about peers’ intent to major in a STEM subject in college or their interest in STEM. Instead, the items available queried about whether peers (1) get good grades, (2) are interested in school, (3) attend classes regularly, and (4) have plans to go to college.

### ***High School STEM Coursetaking and Achievement***

High school coursetaking in STEM related subjects is both affected by and further develops students’ interests in science, technology, engineering, and mathematics. The same can be said about the roles of STEM-related self-efficacy and identity—they can both direct the taking of STEM coursework and as a consequence of successfully completing it, further develop feelings of competence about doing STEM related activities which in turn reinforce science, engineering and mathematics identities. And the stronger the identities, the greater the likelihood of pursuing a STEM major in college.

Previous studies showed that taking high-level coursework in both mathematics and science during high school is key not only to students entering a STEM major, but also to students’ successful completion of a STEM college degree (Tyson et al., 2007; Rask, 2010; Wang, 2013; Gottfried, 2015). The current study found that earning more credits in AP science, engineering and STEM related courses in high school was significantly related to students’ choosing STEM a major in college. In addition, earning at least one credit in Physics during high school was another significant STEM coursetaking factor that increased the likelihood of choosing a STEM major in college even taking into consideration other STEM coursetaking variables like AP science and overall credits in STEM. However, this was not the case for earning at least one credit in Biology or Chemistry. This latter finding was also found by Tyson and his colleague’s study in 2017. They instead speculated that physics course-taking is a primary factor in STEM degree attainment. The current study revealed that taking Physics in high school significantly predicts a students’ chance of choosing a STEM major in college, the prerequisite step to completing a STEM degree. The National Alliance of Black School Educators (2012) stated that

“Physics is a gateway course for post-secondary study in science, medicine, and engineering, as well as an essential component in the formation of students’ scientific literacy [...] College success for virtually all science, computing, engineering, and premedical majors depends in part on passing physics”<sup>9</sup>. However, only 36 percent high school graduates earned at least one Carnegie credit in Physics based on the latest High School Transcript Study data (U.S. Department of Education, 2013). The low percentage of high school students earning credits in Physics might be partially due to the fact that many high schools do not provide Physics courses. Based on the 2015–16 Civil Rights Data Collection, approximately 40 percent of U.S. high schools did not offer Physics. Increased students’ access to higher level STEM courses like Physics might be an effective route to broaden access to the STEM pipeline.

The current study also found that taking applied STEM courses such as engineering is positively associated with the likelihood of students choosing a STEM major in college. Many previous studies found that selection and completion of math and science courses during high school are essential in moving more students toward choosing a STEM major in college (Blickenstaff, 2005). The current study showed that not only the core STEM courses in math and science but also applied courses such as engineering play an important role in engaging more students in STEM fields. Students early exposure to applied STEM courses tends to boost their interest in and competency to master STEM subjects, which in turn serve to motivate them to choose a STEM major in college.

Besides taking STEM courses, doing well in those courses is an important precursor of deciding on a STEM major. The current study included two measures of STEM achievement—students’ STEM GPA and grade 12 NAEP mathematics scores. Both showed significant, substantial relationships with students’ college STEM major enrollment.

### ***Student’s social context***

Students’ family and school contexts are important for understanding their academic decision of choosing a college major. The current study shows that student’s gender and race/ethnicity are significantly associated with the likelihood of students choosing a STEM major in college after taking all other variables into consideration. Female students were less likely to choose a STEM major in college compared to male students. The underrepresentation of women in STEM fields has been fairly persistent over the past decade, even though women have comprised the majority of postsecondary students since the 1980s. The most recent data from the Integrated Postsecondary Education Data System (IPEDS) still show that around 68 percent of 2017–18 STEM degree recipients were male (Snyder et al., 2019). Additional research needs to be conducted to investigate whether female and male students have the same level of STEM motivation and whether the decision process of having a STEM major is impacted differently by motivation and other variables between female and male students.

Regarding student’s race/ethnicity background, although some research has documented that certain race/ethnicity groups are less likely to entering a STEM field in college (Ma & Liu,

---

<sup>9</sup> The position statement is found at <http://vector.nsbp.org/2012/03/16/national-alliance-of-black-school-educators-endorses-physics-first>.



2017), the current study found that the probability of Black or Hispanic students choosing a STEM major at college is not significantly lower compared to White students, after taking into consideration student's STEM motivation, STEM coursetaking and achievement, and their demographic background. The finding might suggest that Black and Hispanic students would have the same likelihood to choose a STEM major if they were sharing the same level of STEM motivation and taking similar STEM courses.

To summarize, the findings from this study add to the extant research literature showing the importance of STEM coursetaking in high school as a predictor of majoring in a STEM subject in college. But perhaps most importantly the results show the central role that motivation plays in who decides to major in a STEM field in a 4-year college. Of note is the role that identities play. Thinking and believing in oneself as a scientist and as a person who is good at mathematics were shown to be key in the choice of a STEM major. Although the research literature suggested that self-efficacy is the key motivational construct for understanding who chooses a college major in a STEM subject, in the presence of measures of science and mathematics identity, self-efficacy showed no direct effect on choosing a STEM major. While this result may seem counterintuitive, perhaps it should not be. To feel efficacious about doing science or mathematics requires that one not only has the ability to solve problems in those arenas but has the confidence to do so and a track record to back up that confidence. An identity in science or mathematics requires that one feels efficacious about solving problems in one or both of these arenas. What distinguishes an identity is that the self—who we see ourselves as being—is comprised of our identities. According to McCall and Simmons (1978) and Stryker (1968) the self can be thought of as a hierarchy of role identities. The more prominent (important) the identity, the more core it is for the self, and the more likely we are to activate it in any given situation. It is precisely because role identities define who we are that they have such strong motivating power. It is for this reason that identities are more important in predicting role-related outcomes than self-efficacy. Self-efficacy is necessary by buttressing and supporting our role identities.

### **Next Steps**

There are several directions in which further research can go. This study revealed that there is a gender gap for students' entrance to a STEM major in college, not only in the observed percentages of the STEM major choices of students of different gender which are well known, but also in the probability of choosing a STEM major while accounting for many other highly relevant predictors in the model. Female students were less likely to claim a STEM major in college despite their equal competences of STEM achievement compared to male students. For example, the most recent releases of grade 8 NAEP mathematics data show that female students have the same scores as male students within a point or two of males (National Report Card, 2019). So why are females less likely to choose a STEM major? Future research could focus on establishing a multi-group SEM model to distinguish how motivation, STEM coursetaking and STEM achievement impact the likelihood of entering a STEM major differently between male and female students. This analysis would enable us to test whether the significant relationships identified in the current study function differently based on students' gender.

A study of gender differences will also be buttressed with the release of HSLS:09 college transcript data in the summer of 2020. This data will allow an exploration of the coursetaking and STEM achievement of females and males in college. Most important will be a determination

of whether females are less likely to take college-level STEM courses than males. If not, can the differences be explained, at least partially, because of differences in females' compared to males' self-efficacy and identity?

Finally, the next iteration of data will be collected when the HSL:09 participants are 30 years old, and when this occurs, it will be important to determine how many of those who majored in STEM subjects are working in STEM occupations. It will be especially important to determine if there are discrepancies by race/ethnicity and gender in the number of STEM majors working in STEM fields as well as between who majors in STEM and who is working in a STEM occupation for gender and or race/ethnicity.

## References

- Abrams, D., & Hogg, M. A. (1988). Comments on the motivational status of self-esteem in social identity and intergroup discrimination. *European Journal of Social Psychology, 18*(4), 317–334.
- Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2013). ‘Not girly, not sexy, not glamorous’: Primary school girls’ and parents’ constructions of science aspirations. *Pedagogy, Culture & Society, 21*(1), 171–194.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191.
- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology, 4*(3), 359–373.
- Bandura, A. (1994). Self-efficacy. In V.S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (vol. 4, pp. 71–81). New York: Academic Press.
- Bandura, A. (1997). *Self-Efficacy: The exercise of control*. New York, NY: W.H. Freeman.
- Beal, S. J., & Crockett, L. J. (2010). Adolescents’ occupational and educational aspirations and expectations: Links to high school activities and adult educational attainment. *Developmental Psychology, 46*(1), 258.
- Benner, A. D., & Mistry, R. S. (2007). Congruence of mother and teacher educational expectations and low-income youth's academic competence. *Journal of Educational Psychology, 99*(1), 140.
- Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in math and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology, 96*(1), 97.
- Bohrnstedt, G.W., Zhang, J., Park, B.J., Ikoma, S., Broer, M. & Ogut, B. (2020). Mathematics identity, self-efficacy, interest and their relationships to mathematics achievement: A longitudinal analysis. In Serpe, R, Stryker, R. & Powell, B. *Identity and Symbolic Interaction: Deepening Foundations; Building Bridges* (pp. 169–210). New York: Springer.
- Bozick, R., Alexander, K., Entwisle, D., Dauber, S., & Kerr, K. (2010). Framing the future: Revisiting the place of educational expectations in status attainment. *Social Forces, 88*(5), 2027–2052.
- Brasier, T. G. (2008). *The effects of parental involvement on students’ eighth and tenth grade college aspirations: A comparative analysis* (dissertation).
- Brenner, P. S., Serpe, R. T., & Stryker, S. (2018). Role-specific self-efficacy as precedent and product of the identity model. *Sociological Perspectives, 61*(1), 57–80.

- Britner, S. L. (2008). Motivation in high school science students: A comparison of gender differences in life, physical, and earth science classes. *Journal of Research in Science Teaching*, 45(8), 955–970.
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, 43(5), 485–499.
- Burke, P. J., & Tully, J. C. (1977). The measurement of role identity. *Social Forces*, 55(4), 881–897.
- Buuren, S. V., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3) 167.
- Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika*, 75(1), 33–57.
- Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, 44(8), 1187–1218.
- Cass, C. A., Hazari, Z., Cribbs, J., Sadler, P. M., & Sonnert, G. (2011). Examining the impact of mathematics identity on the choice of engineering careers for male and female students. In *2011 Frontiers in Education Conference (FIE)* (pp. F2H-1). IEEE.
- Catsambis, S. (2001). Expanding knowledge of parental involvement in children's secondary education: Connections with high school seniors' academic success. *Social Psychology of Education*, 5(2), 149–177.
- Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues*, 67(3), 469–491.
- Cohen, S. M., & Chang, M. (2020). Science achievement within the United States: a view through affective and demographic lenses. *Educational Studies*, 46(2), 221–239.
- Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum.
- Dubina, K. S., Morisi, T. L., Rieley, M., & Wagoner, A. B. (2019). Projections overview and highlights, 2018-28. *Monthly Labor Review*, 142, 1–26.
- Eccles, J. S. (1994). Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly*, 18(4), 585–609.
- Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. *Handbook of Competence and Motivation*, 105–121.

- 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.
- Eccles, 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.
- Eccles, J. S., & Barber, B. L. (1999). Student council, volunteering, basketball, or marching band: What kind of extracurricular involvement matters? *Journal of Adolescent Research*, 14(1), 10–43.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, 21(3), 215–225.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132.
- Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self-and task perceptions during elementary school. *Child Development*, 64(3), 830–847.
- Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). *Motivation to succeed*. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (p. 1017–1095). John Wiley & Sons, Inc.
- Ervin, L. H., & Stryker, S. (2001). Theorizing the relationship between self-esteem and identity. In T. Owens, S. Stryker, & N. Goodman (Eds.), *Extending Self-Esteem Theory and Research: Sociological and Psychological Currents* (pp. 29–55). Cambridge: Cambridge University Press.
- Federman, M. (2007). State graduation requirements, high school course taking, and choosing a technical college major. *The BE Journal of Economic Analysis & Policy*, 7(1).
- Fehrman, P. G., Keith, T. Z., & Reimers, T. M. (1987). Home influence on school learning: Direct and indirect effects of parental involvement on high school grades. *The Journal of Educational Research*, 80(6), 330–337.
- Gasson, R. M., Haller, A. O., & Sewell, W. H. (1972). *Attitudes and Facilitation in the Attainment of Status*. Washington, DC: American Sociological Association.
- Groothuis-Oudshoorn, K., & Van Buuren, S. (2011). Mice: multivariate imputation by chained equations in R. *J Stat Softw*, 45(3), 1–67.
- Gottfried, M. A. (2015). The influence of applied STEM coursetaking on advanced mathematics and science coursetaking. *The Journal of Educational Research*, 108(5), 382–399.

- Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Counseling Psychology, 32*(1), 47–56.
- Hackett, G., & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education, 20*(3), 261–273.
- Hossler, D., & Stage, F. K. (1992). Family and high school experience influences on the postsecondary educational plans of ninth-grade students. *American Educational Research Journal, 29*(2), 425–451.
- Jenson, R. J., Petri, A. N., Day, A. D., Truman, K. Z., & Duffy, K. (2011). Perceptions of self-efficacy among STEM students with disabilities. *Journal of Postsecondary Education and Disability, 24*(4), 269–283.
- Kuenzi, J. J., Matthews, C. M., & Mangan, B. F. (2006). Science, technology, engineering, and mathematics (STEM) education issues and legislative options. Library of Congress Washington DC Congressional Research Service.
- Lameva, B., & Chonteva, Z. (2013). *A cross-ethnic comparison of mathematics achievement in the trends in international mathematics and science study (2011)*. Paper presented at IEA International Research Conference, Singapore.
- Lee, S. W., Min, S., & Mamerow, G. P. (2015). Pygmalion in the classroom and the home: Expectation's role in the pipeline to STEMM. *Teachers College Record, 117*(9), 1–40.
- Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology, 33*(3), 265–269.
- Lent, R. W., Sheu, H. B., Gloster, C. S., & Wilkins, G. (2010). Longitudinal test of the social cognitive model of choice in engineering students at historically Black universities. *Journal of Vocational Behavior, 76*(3), 387–394.
- Leppel, K., Williams, M. L., & Waldauer, C. (2001). The impact of parental occupation and socioeconomic status on choice of college major. *Journal of Family and Economic Issues, 22*(4), 373–394.
- Lichtenberger, E., & George-Jackson, C. (2013). Predicting high school students' interest in majoring in a STEM field: Insight into high school students' postsecondary plans. *Journal of Career and Technical Education, 28*(1), 19–38.
- Little, R. J. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics, 6*(3), 287–296.

- Ma, X. (2001). Participation in advanced mathematics: Do expectation and influence of students, peers, teachers, and parents matter? *Contemporary Educational Psychology*, 26(1), 132–146.
- Ma, Y., & Liu, Y. (2017). Entry and degree attainment in STEM: The intersection of gender and race/ethnicity. *Social Sciences*, 6(3), 89. <https://doi.org/10.3390/socsci6030089>.
- MacPhee, D., Farro, S., & Canetto, S. S. (2013). Academic self-efficacy and performance of underrepresented STEM majors: Gender, ethnic, and social class patterns. *Analyses of Social Issues and Public Policy*, 13(1), 347–369.
- Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science education*, 95(5), 877–907.
- Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review*, 2(2), 77–172.
- Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*, 30(4), 841–860.
- Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80(3), 366.
- McCall, G. J., & Simmons, J. L. (1978). *Identities and interactions: An examination of human associations in everyday life* (Rev. Ed.). New York.
- Moakler Jr, M. W., & Kim, M. M. (2014). College major choice in STEM: Revisiting confidence and demographic factors. *The Career Development Quarterly*, 62(2), 128–142.
- Muthén, L. K., & Muthén, B. (2018). Mplus. *The comprehensive modelling program for applied researchers: user's guide*, 5.
- National Academy of Engineering and National Research Council (2014). *STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/18612>.
- Nurmi, J. E. (2004). Socialization and self-development: Channeling, selection, adjustment, and reflection.
- Ogut, B., Bohrnstedt, G., & Broer, M. (2015). *Imputing 12th-grade NAEP mathematics scores for the full HSLs sample*. Washington, DC: American Institutes for Research.
- Ou, S. R., & Reynolds, A. J. (2008). Predictors of educational attainment in the Chicago longitudinal study. *School Psychology Quarterly*, 23(2), 199–229.

- Pajares, F. (1996). Self-efficacy. (2008). Overview: Self-efficacy in STEM. *SWE-AWE CASEE Overviews*, 1–12.
- Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A Path Analysis. *Journal of Educational Psychology*, 86(2), 193–203.
- Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: the need for specificity of assessment. *Journal of Counseling Psychology*, 42(2), 190–198. <https://doi.org/10.1037/0022-0167.42.2.190>.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in Education: Theory, Research, and Applications* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Raabe, I. J., Boda, Z., & Stadtfeld, C. (2019). The social pipeline: how friend influence and peer exposure widen the STEM gender gap. *Sociology of Education*, 92(2), 105–123.
- Radunzel, J., Mattern, K., & Westrick, P. (2017). Who will declare a STEM major? The role of achievement and interests. ACT Research Report Series 2017-4. *ACT, Inc.*
- Rask, K. (2010). Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review*, 29(6), 892–900.
- Riegle-Crumb, C., Farkas, G., & Muller, C. (2006). The Role of gender and friendship in advanced course taking. *Sociology of Education*, 79(3), 206–228. <https://doi.org/10.1177/003804070607900302>.
- Rittmayer, A. D., & Beier, M. E. Lent, R. W., Lopez, F. G., & Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of Counseling Psychology*, 38(4), 424–430.
- Rutchick, A. M., Smyth, J. M., Lopoo, L. M., & Dusek, J. B. (2009). Great expectations: The biasing effects of reported child behavior problems on educational expectancies and subsequent academic achievement. *Journal of Social and Clinical Psychology*, 28(3), 392–413.
- Sahin, A., Ekmekci, A., & Waxman, H. C. (2017). The Relationships among high school STEM learning experiences, expectations, and mathematics and science efficacy and the likelihood of majoring in STEM in college. *International Journal of Science Education*, 39(11), 1549–1572.
- Sewell, W. H., & Shah, V. P. (1968). Parents' education and children's educational aspirations and achievements. *American Sociological Review* (33), 191–209.
- Sewell, W. H., Haller, A. O., & Ohlendorf, G. W. (1970). The educational and early occupational status attainment process: Replication and revision. *American Sociological Review*, 35(6), 1014–1027. <https://www.doi.org/10.2307/2093379>.



- Snyder, T. D., De Brey, C., & Dillow, S. A. (2019). Digest of Education Statistics 2017, NCES 2018-070. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Stets, J. E., & Burke, P. J. (2000). Identity theory and social identity theory. *Social Psychology Quarterly*, 63(3), 224–237.
- Stets, J. E., Brenner, P. S., Burke, P. J., & Serpe, R. T. (2017). The science identity and entering a science occupation. *Social Science Research*, 64, 1–14.
- Stone, G. P. (1962). Appearance and the self. In A. M. Rose (Ed.), *Human Behavior and Social Processes: An Interactionist Approach* (pp. 86–118). New York: Houghton Mifflin.
- Stryker, S. (1968). Identity salience and role performance: The relevance of symbolic interaction theory for family research. *Journal of Marriage and the Family*, 30(4), 558–564. <https://doi.org/10.2307/349494>.
- Stryker, S. (2003). Whither symbolic interaction? Reflections on a personal odyssey. *Symbolic Interaction*, 26(1), 95–109.
- Stryker, S. (2004). Integrating emotion into identity theory. *Theory and Research on Human Emotions*. Emerald Group Publishing Limited. 1–23.
- Stryker, S., & Burke, P. J. (2000). The past, present, and future of an identity theory. *Social Psychology Quarterly*, 63(4), 284–297.
- Sewell, W. H., & Shah, V. P. (1968). Social class, parental encouragement, and educational aspirations. *American journal of Sociology*, 73(5), 559–572.
- Syed, M., Azmitia, M., & Cooper, C. R. (2011). Identity and academic success among underrepresented ethnic minorities: An interdisciplinary review and integration. *Journal of Social Issues*, 67(3), 442–468. <https://doi:10.1111/j.1540-4560.2011.01709.x>.
- Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Life sci*, 1(0.2).
- Thoits, P. A. (1986). Social support as coping assistance. *Journal of Consulting and Clinical Psychology*, 54(4), 416.
- Trusty, J. (2002). Effects of high school course-taking and other variables on choice of science and mathematics college majors. *Journal of Counseling & Development*, 80(4), 464–474.
- Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, 12(3), 243–270.

- Valian, V. (2007). *Women at the Top in Science--And Elsewhere*. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (p. 27–37). American Psychological Association. <https://doi.org/10.1037/11546-002>.
- Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33*(4), 304–340.
- Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal, 50*(5), 1081–1121.
- Wenger, E. (1998). *Communities of practice: learning, meaning and identity*. New York: Cambridge University Press.
- White, S., & Tyler, J. (2015). Under-represented minorities in high school physics: Results from the 2012–13 nationwide survey of high school physics teachers. *Focus On*.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology, 25*(1), 68–81.
- Wigfield, A., Eccles, J. S., Davis-Kean, P., Roeser, R., & Scheifele, U. (2006). Motivation to succeed. *Handbook of Child Psychology. Social, Emotional, and Personality Development, 3*, 933–1002.
- Williams, M. M., & George-Jackson, C. (2014). Using and doing science: Gender, self-efficacy, and science identity of undergraduate students in STEM. *Journal of Women and Minorities in Science and Engineering, 20*(2).
- U.S. Department of Education, National Center for Education Statistics. (2013). Table 223.40. Percentage of public and private high school graduates taking selected mathematics and science courses in high school, by selected student and school characteristics: Selected years, 1990 through 2009. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of Education Statistics* (2013 ed.). Retrieved from [https://nces.ed.gov/programs/digest/d13/tables/dt13\\_225.40.asp](https://nces.ed.gov/programs/digest/d13/tables/dt13_225.40.asp).
- Zhang, J., Bohrnstedt, G., Park, B.J., Ikoma, S., Ogut, B., Broer, M. (2018). *Mathematics motivation and its relationship with mathematics performance: evidence from the NAEP-HSLS overlap sample*. (Research report for NCES). Washington, DC: American Institutes for Research.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91. <https://doi:10.1006/ceps.1999.1016>.
- Zimmerman, B. J., & Kitsantas, A. (1997). Developmental phases in self-regulation: Shifting from process goals to outcome goals. *Journal of educational psychology, 89*(1), 29.

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of educational Psychology, 91*(2), 241.

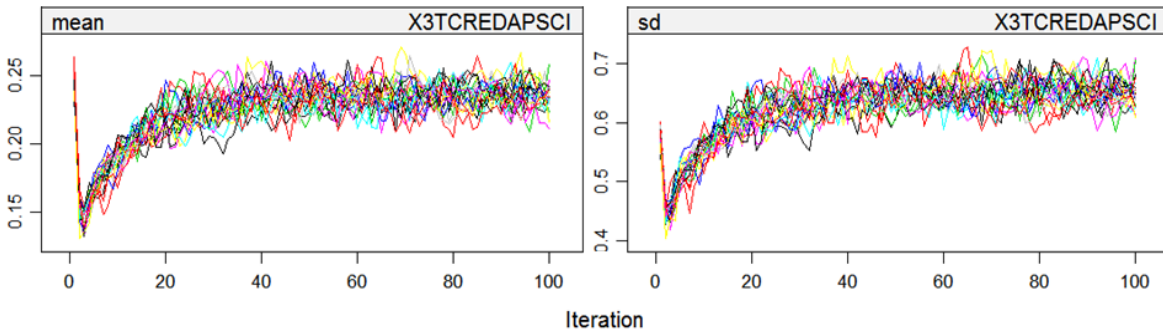
## Appendix A: Nested Imputations of Projected HSLS NAEP Scores

The projected NAEP mathematics achievement scores for HSLS:09 participants were imputed in two stages, namely multiple imputations of HSLS:09 background variables and plausible value generation in a latent regression model.

In the first stage, multivariate imputations by Chained Equations (MICE; van Buuren & Groothuis-Oudshoorn, 2011) method was used to impute the missing values for both level-1 student variables as well as level-2 school variables in HSLS:09. The method started with a multi-level regression model to provide predicted values of missing data for a variable, given observed data. Once this variable was imputed, it was used in a sequence to impute the next variable. The sequence is repeated multiple times until convergence. The random draws of imputed values were implemented via predictive mean matching (PMM; Little, 1988), a semi-parametric imputation method that limits imputations to observable values only. A key advantage of PMM is that it can preserve non-linear relations even if the imputation model is wrong (MICE; Groothuis-Oudshoorn & van Buuren, 2011). In our study, 100 iterations were conducted for 20 sets of imputations. The imputation diagnostics were performed by conducting convergence checks. Convergence was decided by examining the trace-plots of multiple sets of the imputed values over the number of iterations. An example of the trace-plot for the variable X3TCREDAPSCI (the number of credits earned in AP science) is shown in Figure A-1. As can be seen in Figure A-1, the mean and the standard deviation of the variable for all 20 chains stabilize after the 40<sup>th</sup> iteration approximately, which is an indicator of convergence.

In the second step, plausible values of students' NAEP mathematics scores were generated in a latent regression model for each set of imputed background variables. Specifically, a five-dimensional IRT model (representing five NAEP mathematics subscales) was fitted to the HSLS:09 sample's NAEP response data, where the non-overlap students' responses were treated as missing at random. The item parameters were fixed at operationally calibrated values in NAEP. The five subscales were then regressed on each set of the imputed background variables separately via a Metropolis-Hastings Robins-Monro (MH-RM; Cai, 2010) algorithm in flexMIRT software. Twenty sets of plausible values were randomly drawn for the MH-RM run. Finally, the plausible values were linearly transformed to NAEP reporting scale using operational NAEP transformation coefficients. The latent regression model was repeated for each set of imputed background variables from the first stage. In total, 400 sets of plausible values (20 sets of plausible values for each of the 20 sets of imputed background variables) for NAEP mathematics scores were generated for the HSLS:09 sample.

Figure A-1. Examples of trace-plots for means and standard deviations of one variable over iterations for MICE convergence checks.



NOTE: X3TCREDAPSCI refers to the number of credits earned in AP science.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLSL:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

As a quality check, the means and standard deviations of the imputed plausible values of the overlap sample were compared to those from NAEP to see if they have a similar marginal distribution. The means and standard deviations of the pooled NAEP operational plausible values and two examples of the 400 sets of plausible values imputed in this study are presented in Table A-1 and Table A-2. The comparisons indicated that there is little difference between our imputations and the operational ones for the overlap sample across subscales and the composite scale.

Table A-1. Means of operational NAEP plausible values and imputed plausible values for the overlap sample

	Number	Measurement	Data	Algebra	Composite
Operational NAEP PV (pooled)	152.39	156.26	156.59	158.86	156.86
First PV of first imputation	154.32	156.53	156.57	158.45	156.99
First PV of second imputation	154.56	156.61	157.15	158.55	157.22

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLSL:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

Table A-2. Standard deviations of operational NAEP plausible values and imputed plausible values for the overlap sample

	Number	Measurement	Data	Algebra	Composite
Operational NAEP PV (pooled)	33.28	31.14	35.64	32.75	31.76
First PV of first imputation	31.33	31.12	37.94	33.04	32.10
First PV of second imputation	31.24	31.12	37.77	33.35	32.14

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLSL:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

Besides the means and variances, we also compared the correlations between the subscales for the overlap sample between our imputations and the operational NAEP results. Table A-3 presents the correlations between subscales and the composite scales using operational plausible values, while those using our own imputations are presented in Table A-4. The pairwise correlations are almost identical between the two sets of plausible values.

Table A-3. Correlations of subscales and the composite scale using operational NAEP plausible values for overlap sample

NAEP subscale or scale	Number	Measurement	Data	Algebra	Composite
Number	1				
Measurement	0.87	1			
Data	0.86	0.90	1		
Algebra	0.87	0.91	0.89	1	
Composite	0.92	0.97	0.96	0.97	1

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

Table A-4. Correlations of subscales and the composite scale using nested-imputation NAEP plausible values for overlap sample

NAEP subscale or scale	Number	Measurement	Data	Algebra	Composite
Number	1				
Measurement	0.87	1			
Data	0.87	0.89	1		
Algebra	0.86	0.90	0.88	1	
Composite	0.92	0.96	0.96	0.97	1

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

## Appendix B

Table B-1 presents intercorrelations among the latent motivation variables (including both mathematics and science motivation), high school STEM coursetaking variables, STEM achievement variables, social network variables and STEM major.

Table B-1. Correlations between mathematics motivation, science motivation, STEM coursetaking, background information, and STEM major

	[a]	[b]	[c]	[d]	[e]	[f]	[g]	[h]	[i]	[j]	[k]	[l]	[m]	[n]	[o]	[p]	[q]	[r]	[s]	[t]
a. STEM major	1																			
b. Mathematics identity	0.33	1																		
c. Mathematics self-efficacy	0.27	0.67	1																	
d. Science identity	0.33	0.29	0.24	1																
e. Science self-efficacy	0.20	0.20	0.35	0.56	1															
f. Educational expectations	0.14	0.17	0.20	0.27	0.21	1														
g. NAEP Score	0.29	0.47	0.29	0.28	0.19	0.22	1													
h. Credits AP math	0.24	0.33	0.20	0.19	0.13	0.18	0.53	1												
i. Credits AP science	0.28	0.23	0.19	0.27	0.11	0.23	0.44	0.52	1											
j. $\geq 1$ credit physics	0.20	0.24	0.18	0.15	0.03	0.09	0.33	0.24	0.22	1										
k. Credits engineering	0.16	0.15	0.11	0.07	0.05	0.02	0.08	0.04	0.03	0.09	1									
l. Credits STEM	0.26	0.25	0.23	0.23	0.12	0.11	0.20	0.27	0.30	0.31	0.35	1								
m. GPA STEM	0.22	0.40	0.28	0.21	0.17	0.24	0.48	0.31	0.27	0.16	0.06	0.26	1							
n. Parental STEM occupation	0.07	0.06	0.03	0.08	0.04	0.02	0.15	0.11	0.11	0.08	0.01	0.06	0.06	1						
o. Parental educational expectations	0.10	0.11	0.12	0.08	0.07	0.23	0.17	0.13	0.17	0.09	0.00	0.10	0.17	0.01	1					
p. Friends' influences	0.04	0.02	0.05	0.04	0.03	0.13	0.10	0.12	0.17	0.06	0.04	0.14	0.19	0.08	0.11	1				
q. Family SES	0.09	0.06	0.04	0.13	0.05	0.13	0.32	0.18	0.20	0.14	0.01	0.07	0.24	0.25	0.14	0.17	1			
r. School percentage FRPL eligible students	0.05	0.02	0.02	0.03	0.06	0.01	0.26	0.08	0.07	0.12	0.00	0.02	0.12	0.12	0.05	0.07	0.41	1		
s. School percentage of EL students	0.02	0.00	0.02	0.02	0.01	0.02	0.02	0.12	0.08	0.02	0.06	0.08	0.08	0.01	0.02	0.06	0.14	0.31	1	
t. School percentage of SD students	0.01	0.01	0.00	0.02	0.00	0.06	0.09	0.04	0.04	0.02	0.05	0.05	0.01	0.01	0.06	0.01	0.11	0.29	0.04	1

NOTE: [a] STEM major refers to STEM majors at a 4-year college. [h] Math is an abbreviation for mathematics. [r] FRLP refers to Free/Reduced Priced Lunch Program. [s] EL refers to English learners. [t] SD students refers to students with disabilities.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.



## Appendix C

Table C-1 presents the detailed information regarding the path coefficients and the disturbance term of the SEM model.

Table C-1. Coefficient descriptions

Coefficients	Representation
$Y_i$	Choice of STEM major
$\beta_0$	Intercept
Motivation	
$\beta_1$	Mathematics identity
$\beta_2$	Mathematics self-efficacy
$\beta_3$	Science identity
$\beta_4$	Science self-efficacy
$\beta_5$	Educational expectations
STEM coursetaking	
$\beta_6$	Credits AP math
$\beta_7$	Credits AP science
$\beta_8$	At least 1 credit physics
$\beta_9$	Credits engineering
$\beta_{10}$	Credits STEM
STEM achievement	
$\beta_{11}$	NAEP mathematics score
$\beta_{12}$	STEM GPA
Social network	
$\beta_{13}$	Parental STEM occupation
$\beta_{14}$	Parental educational expectations
$\beta_{15}$	Friends' influences
Demographic background	
$\beta_{16}$	Female
$\beta_{17}$	Black
$\beta_{18}$	Hispanic
$\beta_{19}$	Asian
$\beta_{20}$	Other
$\beta_{21}$	Family SES
School background	
$\beta_{22}$	Percentage of students with free/reduced price lunch
$\beta_{23}$	Percentage of EL students
$\beta_{24}$	Percentage of students with disabilities
$\varepsilon_i$	Disturbance term (uncorrelated with any of the independent variables)

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLs:09), First Follow-up (2012), 2013 Update, High School Transcripts (2013–2014), and Second Follow-up (2016); 2013 National Assessment of Educational Progress (NAEP), Grade 12 Mathematics Assessment.

*This page was intentionally left blank.*

## About the American Institutes for Research

Established in 1946, with headquarters in Arlington, Virginia, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance to solve some of the most urgent challenges in the U.S. and around the world. We advance evidence in the areas of education, health, the workforce, human services, and international development to create a better, more equitable world. The AIR family of organizations now includes IMPAQ, Maher & Maher, and Kimetrica. For more information, visit [AIR.ORG](https://www.air.org).



**American Institutes for Research®**  
1400 Crystal Drive, 10th Floor  
Arlington, VA 22202-3289  
202.403.5000 | [AIR.ORG](https://www.air.org)