

RESEARCH

Open Access



Perceived abilities or academic interests? Longitudinal high school science and mathematics effects on postsecondary STEM outcomes by gender and race

Teng Zhao¹ and Lara Perez-Felkner^{2*}

Abstract

Purpose of the study: Previous literature has examined the relationship between high school students' postsecondary STEM major choices and their prior interest and perceived ability in mathematics. Yet, we have limited understanding of whether and how perceived ability and interest in science and mathematics jointly affect students' STEM major choices.

Results: Using the most recent nationally representative longitudinal cohort of U.S. secondary school students, we examine the degree to which students' perceived mathematical and scientific abilities and interests predict their STEM major choices, employing logistic regression and a series of interaction analyses. We find that while both mathematics and science perceived ability positively influence STEM major selection, academic interest in these subjects is a weaker predictor. Moreover, across a series of analyses, we observe a significant gender gap—whereby women are less than half as likely to select STEM majors—as well as nuanced distinctions by self-identified race. The relationships among perceived ability, interest, and STEM major choice are not found to meaningfully vary by race nor consistently by gender. However, perceived ability has a more positive effect for men than women who are pursuing Computing/Engineering majors and a more positive effect for women than men who are pursuing other STEM majors, including less applied Social/Behavioral, Natural, and Other Sciences.

Implications: These findings suggest potential opportunities to enhance their perceived mathematical and scientific abilities in high school, positioning them to potentially enter STEM fields. School sites with more resources to support the ambitions of STEM students of all backgrounds may be better positioned to reduce postsecondary disparities in STEM fields. Given existing opportunity gaps and resource differentials among schools, corresponding recommendations are suggested.

Keywords: Science education, Perceived ability, Academic interest, STEM education, Postsecondary majors

Introduction

Science, technology, engineering, and mathematics (STEM) fields are growing in demand and pay twice as much as non-STEM occupations (Fayer et al., 2017). Accordingly, U.S. high schools offer more advanced mathematics and science opportunities to better position students for postsecondary STEM degrees. Notably, while we have decades of research on mathematical

*Correspondence: lperezfelkner@fsu.edu

² Department of Educational Leadership and Policy Studies and Center for Postsecondary Success, Florida State University, 1114 West Call Street, Tallahassee, FL 32306, USA
Full list of author information is available at the end of the article

ability and interest as predictors of postsecondary STEM outcomes, limited research has directly assessed the interplay and distinctions between these predictors. Even less is known about science ability and interest, after decades of research focused particularly on mathematics.

Early exposure to STEM-related courses can awaken students' interests and lead them to postsecondary STEM majors (Bottia et al., 2015). Advanced mathematics and science courses in high school have been found to be essential to students' opportunities to study STEM subjects in college (Dalton et al., 2007; Schneider et al., 2013). Psychological factors, including interest, have been found to influence students' pathways into these courses (Milesi et al., 2017; Perez-Felkner et al., 2017). Often regarded as a motivational variable and/or an affective component in educational psychology, interest has been widely studied as a mechanism that stimulates students' learning and academic achievement (Renninger & Hidi, 2011; Renninger et al., 1992). However, we know less about how academic interests influence students' postsecondary degree field. Self-assessed mathematics ability has also been found to affect students' STEM major choices, even when students' self-assessments of ability are biased by sociocultural norms around who belongs in these fields (Beyer, 1990; Correll, 2001). Interest is similarly socially conditioned (Watt et al., 2012). Complicating these educational metrics of self-assessed ability and interest in science and mathematics are the complex socioeconomic, racial, and gender dynamics currently at play in secondary and postsecondary education. Socioeconomic inequality by race and gender remains a major social problem, even among college graduates (Doren & Lin, 2019; Pais, 2011). Since STEM degrees can lead to highly paid jobs, one would expect STEM professions to attract underrepresented students and women who are increasingly the primary earners in their family. However, racial and gender disparities continue to persist in these fields (Carter et al., 2019; Saw et al., 2018; Xie et al., 2015). Importantly, gender differentials in post-college earnings appear to be meaningfully explained in part by postsecondary degree field (Bobbitt-Zeher, 2007; Xu, 2015; Zhang, 2008). Whether this is also the case for racial disparities remains unclear.

Given the persistent gender and race disparities in White- and male-dominant STEM fields, despite extensive investment in broadened access and a changing economy, we examine a contemporary cohort comprising a full range of socioeconomic, racial, and gender demographics with a focus on both mathematics and science predictors of STEM postsecondary education. Using the National Center for Education Statistics' most recent longitudinal data, High School Longitudinal Study of 2009 (HSL: 09), this study investigates (1) the distinct effects

of mathematical and scientific interests and ability beliefs on STEM major choice; (2) whether interests or abilities have stronger effects on students' STEM major choice; (3) specific effects by students' self-reported gender and racial identity, and (4) whether these patterns differ in by STEM major cluster, with specific attention to Computing/Engineering majors.

Literature review

Disparities in U.S. postsecondary STEM education and beyond

There are clear gender and racial disparities in STEM postsecondary educational outcomes (e.g., Griffith, 2010; Huang et al., 2000; Shapiro & Sax, 2011). Riegler-Crumb and Peng (2021) found that gender differences in postsecondary major choice can be shaped by societal stereotype and self-beliefs about mathematics ability. Students' pathways to STEM degrees may differ among students by gender and race (Ireland, et al., 2018). Using an intersectional approach, Nix and Perez-Felkner (2019) found that Black women and men experience especially strong gains from positive mathematics ability beliefs and—after controls—are the most likely to declare mathematics-intensive STEM majors and earn degrees in these fields.

Historically, Black, Latina/o/x, and Native American students have been characterized as underrepresented groups in STEM disciplines (Estrada et al., 2016; Maltese & Tai, 2011). Yet, racially minoritized students such as Black and Latina/o/x youth were *more* likely to major in certain STEM disciplines than their White peers when using multivariate analyses to account for explanatory factors such as secondary school course preparation and postsecondary enrollment (Nix & Perez-Felkner, 2019), and racial differences are effectively null when focusing on the population of students who enroll in college (e.g., Riegler-Crumb & King, 2010). This is not the case for gender, where disparities remain among the U.S. postsecondary student population. Such research supports the value of ongoing investigation into the mechanisms which might explain STEM postsecondary major selection, as it varies by gender and race (see Garrison, 2013; Riegler-Crumb, King, et al., 2019; Riegler-Crumb, Morton, et al., 2019; Xu, 2016).

Studies using nationally representative, longitudinal data on U.S. cohorts of students have advanced our understanding of these patterns; these data follow students from secondary school through postsecondary education and workforce entry. In Griffith's (2010) analysis of National Education Longitudinal Study of 1988 (NELS: 88) and National Longitudinal Study of Freshmen data, White and male students comprised higher shares of those persisting in STEM disciplines when compared to underrepresented groups and women. Secondary

STEM course completion can explain much of the race gap in STEM postsecondary outcomes, including STEM major choice (Tyson et al., 2007) and entry into specific mathematics-intensive STEM major clusters (see e.g., Riegle-Crumb & King, 2010; Perez-Felkner et al., 2014) such as Computing and Engineering fields.

These studies did not however assess the effects of science ability beliefs nor closely attend to interest in science. This is now possible with the newest U.S. national cohort: the High School Longitudinal Study of 2009 (HLS: 09), which traces students from secondary school through postsecondary degrees. To reduce demographic inequality of opportunity to enter STEM career fields, it is crucial to understand the factors that may affect students' STEM major choices, which may lead them to a STEM career.

Contextual factors influencing selection of STEM and Computing/Engineering majors

In recent years, a steady flow of research has highlighted factors that may affect students' STEM major selection (Shapiro & Sax, 2011; Wang, 2013a). In addition to the demographic characteristics discussed above, socioeconomic characteristics such as parental occupation and educational background also positively predict postsecondary STEM outcomes (Wagner et al., 2002). For instance, Oguzoglu and Ozbeklik (2016) found in the 1979 National Longitudinal Study of Youth that girls whose fathers who were employed in a STEM profession were more likely to choose a STEM major in college, if they had no male siblings. In addition, family income was one of the components of family socioeconomic status that was used in predicting students' STEM major choices (Niu, 2017). High-demand and high-earning fields such as Computing and Engineering may be especially attractive to students from less socioeconomically advantaged families, as compared to less applied and high-earning scientific fields.

Looking further into specific major fields such as Computing and Engineering vs. applied and non-applied Humanities and Social Science fields, Wiswall and Zafar (2015) found that variations in individual beliefs about ability in each major affect students' major intentions, exacerbating the gender gap in major choices. Different majors tend to have distinct associations with abilities. Focusing on mathematics-intensive majors, Nix and Perez-Felkner (2019) found that 12th grade girls' perceptions of their ability with difficult mathematics increased their likelihood of choosing mathematically intensive majors such as engineering and computer sciences over biology and social/behavioral science majors. Dika and D'Amico (2016) also found that students' perceived preparation in math could significantly predict students'

persistence in these majors. Notably, in the fields of Computing and Engineering, women have remained distinctly underrepresented (Corbett & Hill, 2015). Thus, this study digs further into these fields specifically.

The demographic characteristics of students' schools are also commonly included as covariates in statistical models in prior research that examine students' STEM major choices, such as high school type (e.g., Wang, 2013a) and urbanicity (e.g., Legewie & DiPrete, 2014). Quadlin (2017) used the 1997 National Longitudinal Study of Youth cohort and found that students from less economically advantaged family backgrounds had a higher probability of majoring in applied non-STEM fields such as business, communications, and education disciplines as compared to higher-income families. Generally, private high schools offer more advanced mathematics and science courses and have a higher socioeconomic and college preparatory profile (Lee et al., 1998). As such, students in private schools are better positioned for postsecondary STEM majors and STEM careers (Ketenci et al., 2020). School urbanicity matters as well. Rural students have been found less likely to complete advanced science courses than those in urban schools (Perez-Felkner et al., 2014). Bottia et al. (2018) found that rural high schools were less likely to have STEM-focused programs, which could positively affect students' postsecondary major intentions.

In sum, it seems important to consider these contextual factors which may affect students' pathways into STEM disciplines, during and beyond high school. Wang (2013b) found a stronger effect of math self-efficacy on 2-year college students' STEM interests than on 4-year college students. Phelps et al. (2018) found that students' STEM major enrollments differed at 2-year and 4-year colleges (e.g., mechanical technologies vs. engineering and engineering technology, respectively), within if not between major clusters such as Computing/Engineering fields. Gender differences in U.S. students' attainment of Natural/Engineering Sciences and Life Sciences majors has also been found to vary by college type, after controlling for additional characteristics such as college STEM GPA and academic and social integration (Perez-Felkner et al., 2019). Following these discussions of the value of explaining disparities in postsecondary STEM major selection and contextual explanations, we turn to developmental and social psychological frameworks, which are the focus of our study.

Theorizing postsecondary STEM major choice

Psychological theories have been adapted to explain students' pathways from earlier educational experiences through postsecondary STEM outcomes, including major selection. This manuscript focuses especially on

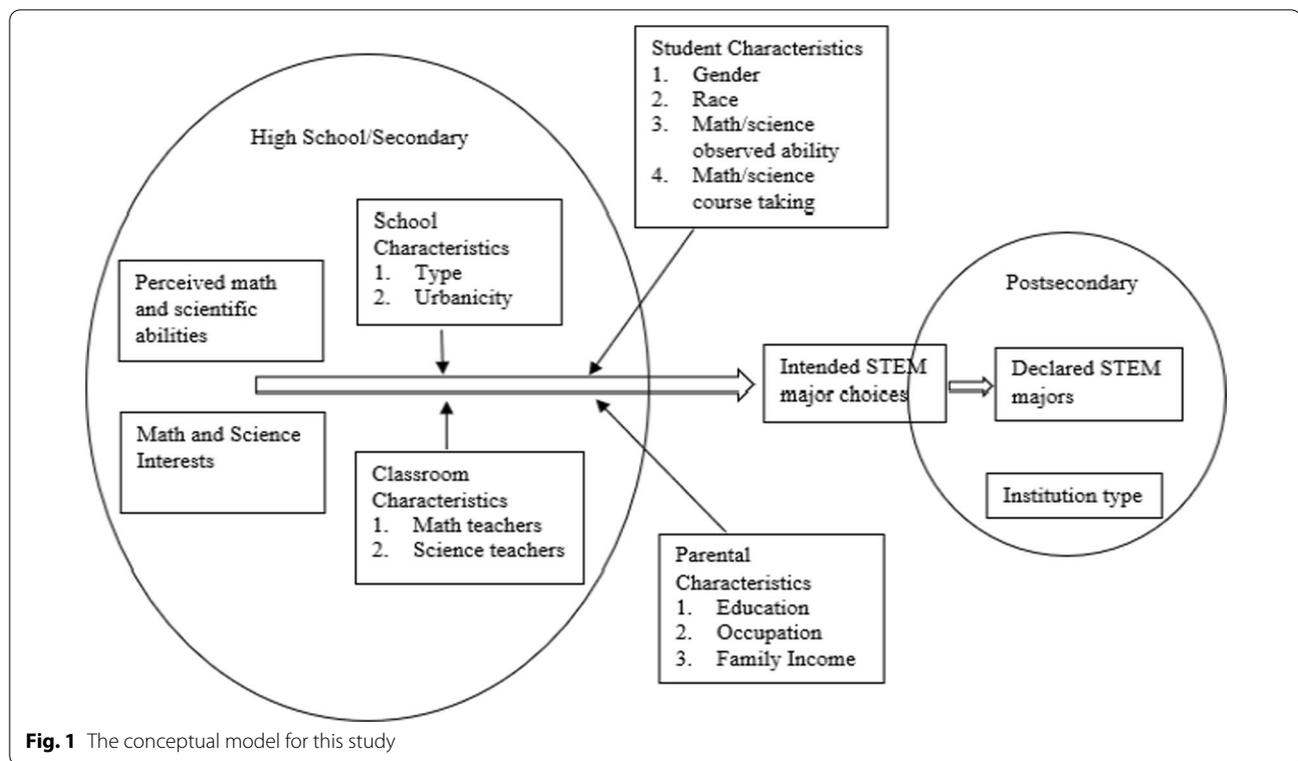


Fig. 1 The conceptual model for this study

students’ perceived abilities and interests in mathematics and science. We also considered complementary theory which attends to the motivational relationships between students’ (1) expectations for success and (2) subjective task values; expectancy task-value theory postulates that individuals’ achievement-related choices (such as post-secondary majors) are associated with their confidence in specific subject domains, such as mathematics and science (Eccles & Wigfield, 2002). Subjective task value has been further categorized into four subcomponents including intrinsic value—the enjoyment obtained from participating in these tasks (Eccles & Wigfield, 2002).

In turn, college students may be more likely to choose the majors in which they feel more efficacious and interested. This may be especially true for students from underrepresented backgrounds who typically receive less encouragement and support in their pursuit of STEM degrees. Using an older cohort of U.S. longitudinal data, Perez-Felkner et al. (2017) also found that perceived mathematical ability affects science course enrollment in secondary school, which in turn has consequences for postsecondary major choice. Measures for science ability were not available in their study. However, students with higher perceived mathematical abilities were found more likely to declare STEM postsecondary majors. Perceived mathematical abilities had particularly positive effects on the probability of Black girls’ and boys’ selection of

mathematics-intensive STEM majors and completing degrees in these fields (Nix & Perez-Felkner, 2019).

There has been more limited attention to how mathematics and science interest may affect students’ STEM major choice, despite academic interest being an important concept in educational psychology (Renninger et al., 1992). Therefore, this study examines whether interest—in conjunction with other established factors—predicts the likelihood of majoring in STEM fields and demographic disparities in these outcomes. This study will also assess students’ perceived ability in mathematics—already shown to affect postsecondary STEM pathways—and science, a crucial but under-examined domain.

Conceptual framework

This study develops a framework which integrates Bandura’s (1977) self-efficacy, Hidi and Renninger’s (2006) interest development model, and Eccles et al. expectancy-value theory (Eccles, 1983; Eccles & Wigfield, 2002), while accounting for prior literature on factors predicting students’ STEM major choices. In this model (see Fig. 1), students’ intent to choose a STEM major is influenced by their perceived mathematical and/or scientific abilities, mathematical and/or scientific interests, parental education and occupations, high school and classroom characteristics, which may stimulate students’

STEM motivation. This study presents a detailed explanation of the model's theoretical grounding.

Ability-related perceptions and beliefs

Postsecondary STEM majors require strong foundational knowledge of mathematics and science. High school advanced mathematics and science courses can signal preparedness to enter college in general and to enter postsecondary gateway courses to STEM majors specifically (Schneider et al., 2013; Tyson et al., 2007). These courses also have implications for students' perceived mathematical ability, including how girls and underrepresented students' self-assessments of their ability may be shaped by in these gendered course and school contexts, with implications for entry to STEM majors (Correll, 2001; Perez-Felkner et al., 2014).

Bandura's (1977) self-efficacy model emphasized an individual's belief in their innate capacity to achieve a particular goal. When individuals realize that their abilities may not enable them to accomplish certain goals, they may give up. The concept of self-efficacy has been widely used in education research. Zimmerman et al. (1992) found that students' perceived academic self-efficacy significantly affects their educational goal setting. Bandura (1993) illustrated that self-efficacy could be functioning in the academic field through four major processes, including motivational and selection processes. Students' beliefs in their efficacy played an important role in regulating their academic activities and aspirations, as well as their motivation directly. Later, Bandura et al. (2001) demonstrated that children's self-efficacy could be more decisive in shaping their occupational decisions than actual academic achievement, which in turn indirectly influences students' major choices. Therefore, self-efficacy is connected to goal setting as well as motivation and/or aspiration, thereby influencing students' STEM-related experiences in secondary school. In turn, building on Bandura's concept of self-efficacy, perceived mathematical and scientific abilities may each shape high school students' pursuit of STEM degrees in college towards the particular goal of majoring in STEM fields. We assess students' perceived abilities holistically as described further in the methodology, encompassing self-assessments of ability on specific tasks (tests, assignments, difficult textbook material) and with their mathematics/science course and the discipline more generally, during their first high school year of study in these crucial subjects for future STEM majors.

Interest development

Hidi and Renninger's (2006) Four-Phase Interest Development Model illustrates how a person's interest influences their attentions, goals, and levels of learning,

which may contribute to academic motivation and/or aspiration. Hidi and Harackiewicz (2000) identify interest as a crucial motivational variable that affects students' academic performance. Thus, if students have stronger interest in mathematics and science, they may be more likely to have better mathematics and science performance and/or ability, potentially influencing their STEM choices.

According to Hidi and Renninger's model, interest is categorized as situational or individual. Situational interest is a temporary interest aroused by specific activities, while individual interest is a relatively stable interest (Schiefele, 2009). Situational interest can develop into individual interest, where a person's subsequent decisions/behaviors are associated with this well-developed interest. Notably, scholars have argued that individual interest is not exclusively generated by the individual (Csikszentmihályi et al., 1993); School actors such as peers and teachers may affect the formation of well-developed individual interest. Frenzel et al. (2010) found students' development of mathematics interest was positively associated with classroom characteristics like mathematics teacher enthusiasm. Therefore, students' classroom experiences with their mathematics and science teachers may also affect their interests in these areas.

Research questions and hypotheses

Guided by prior literature and the conceptual framework above, this study examines the influence of perceived abilities and interest in mathematics and science on students' STEM major choice. Specifically, this study addresses the following four research questions:

1. What is the relationship between postsecondary students' STEM major choice and their perceived abilities and interests in science and mathematics in high school?
2. To what extent do perceived abilities and academic interests in science and mathematics *distinctly* predict students' STEM major choice?
3. To what degree do these relationships vary by gender and/or race?
4. To what extent do these patterns differ in predicting students' choice of Computing/Engineering, Other STEM majors, vs. non-STEM majors?

Students with higher perceived mathematical/scientific abilities may be more likely to choose STEM disciplines in college. Similarly, students who have higher math/science interests are more likely to choose STEM disciplines. It is unclear whether perceived abilities or academic interests are the stronger predictor of STEM

major choice. Research on perceived abilities and academic interest discussed above suggest that both could positively affect students' corresponding choices. Given interest in developing interventions to enhance students' opportunities to enter and complete STEM majors, we are interested in which is a more effective predictor of students' selection of STEM majors. Additionally, we evaluate whether these relationships vary by gender and race. Finally, we consider whether the analytic models are sensitive to alternate specifications of the dependent variable. We assess these relationships on a more nuanced dependent variable, which parses high-growth and high-earning applied majors in technological fields (Computing and Engineering) as compared to other STEM and non-STEM fields.

Methods

Data source, declarations, and participants

We used restricted-use data from the newest nationally representative longitudinal U.S. cohort, the High School Longitudinal Study of 2009 (HSL: 09). The National Center for Education Statistics (NCES) followed incoming ninth-grade students through secondary and postsecondary education, beginning in fall 2009. Students completed follow-up surveys in spring 2012 when the majority were in eleventh grade and at the transitory point from secondary to postsecondary education in 2013, to collect information like college plans and choices after high school completion (Duprey et al., 2018; Ingels et al., 2014). An additional follow-up survey was administered in spring 2016. Beyond these surveys, federal data from other sources was incorporated into the restricted-use HSL: 09 dataset by NCES and is acknowledged where appropriate in the table source information we report in the main paper and appendix. Public use access to these data is available through the NCES website, as is the application for restricted-use data, such as we used for these analyses: https://nces.ed.gov/surveys/hsls09/hsls09_data.asp. Statistical code generated to analyze supporting the findings of this study are available from the corresponding author upon request.

The base-year sample includes ninth graders from 940 high schools across the country. Here and throughout, descriptive information is rounded to the nearest tenth in accordance with NCES restricted-data licensing regulations. Parents, teachers, school administrators, and counselors were also surveyed. The base-year student survey asked questions related to students' perceived mathematical/scientific abilities and math/science interests as well as their feelings about math/science teachers. The base-year parent survey collected information about students' family characteristics. Students' first-declared college major was gathered in the postsecondary follow-up

student survey (see detailed variable descriptions Table 7 in Appendix).

The analytic sample represents the students who were ninth graders in Fall 2009 and were still college students in spring 2016. We restricted the sample to those who attended postsecondary institutions. Because there are missing cases associated with each analytic variable, we used multiple imputation to reduce nonresponse bias. After imputation, our final analytic sample yielded 11,560 cases. To ensure generalizability to the U.S. student population and enhance external validity, we used panel strata and survey weights (*w4w1stup1*) and provided by HSL: 09 to address the complex stratified sampling design, and to adjust for unequal selection probabilities of sub-populations, consistent with the recommendations by Duprey and colleagues (2018) for appropriate weights when analyzing multi-wave HSL: 09 data.

Measures

This section summarizes all the variables used in the hypothesized model (Fig. 1).

STEM major The primary dependent variable in this study represents whether a student chooses a STEM major, a dichotomous variable (1 = first major is STEM; 0 = not STEM). According to the HSL: 09 codebook, the source variable (X4RFDGMJSTEM) represents the respondent's major or field of study as of February 2016. STEM designations follow the U.S. Department of Education's (2010) Classification of Instructional Programs. We also developed and investigated an alternate three-category specification drawn from related source variable X4RFDGMJ14Y, that breaks into STEM major clusters: (1) Computing/Engineering fields, (2) Social/Behavioral, Natural, and Other Sciences, and (3) non-STEM fields. Past research has shown Computing and Engineering function distinctly from other scientific areas given the high earnings and growth associated with these applied and mathematically intensive technological fields (see Corbett & Hill, 2015; Scott, et al., 2015). The non-STEM category is identical across both groups, as indicated later in Table 2 and Table 7 in Appendix.

Perceived mathematical/scientific ability Multiple items in the base-year student survey measured students' perceived ability. We identified seven items that include ability self-assessments and their reflection in their perceptions of themselves as a mathematics/science "person"; this scale is intended to richly and robustly reflect students' perceived mathematical/scientific abilities (see Table 8 in Appendix). Original items were all 4-point Likert scale items, ranging from 1 (strongly agree) to 4 (strongly disagree), which we reverse-coded for interpretability such that the items range from strongly disagree to strongly agree as they increase from 1 to 4. The

Cronbach's reliability α for our perceived mathematical and scientific ability scales are 0.89 and 0.87, respectively, indicating strong internal consistency among these seven items (see Kline, 2013). As an additional robustness check, we assessed the correlations between our original scales for students' perceived mathematical and scientific abilities and the NCES-provided measures of mathematics self-efficacy ($r=0.92, p<0.001$) and science self-efficacy ($r=0.70, p<0.001$).¹ More detailed information about these and subsequently discussed survey items is provided in Table 8.

Math/science interest Similarly, mathematics and science interests were also measured by a series of items (details Table 9 in Appendix). Again, the original score on each item ranged from 1 to 4 (strong agreement) and 4 (strong disagreement). Two negatively worded items were found. We therefore reverse-coded these measures, such that four indicates "strongly agree", to be consistent with high values meaning high interests. The Cronbach's reliability α for the scales we generated to represent mathematics and science interests are 0.78 and 0.81, respectively. These three items thus represent math/science interest reasonably well, as reliability scores of $\alpha > 0.75$ are in the acceptable range of internal consistency (Kline, 2013).

Student characteristics We included students' demographic characteristics, include dichotomous variables coded 1/0 for students' self-reported gender and race/ethnicity: White, Asian, Black, Latina/o/x, and Multiple/Other race. We also account for students' family socioeconomic status, as designated by parental education and family income. Parental education was distinguished by whether at least one parent had a 4-year college degree. Family income is an ordinal scale representing family income from all sources in 2008, the year prior to ninth grade, and is the most detailed family income variable included the HSLs restricted-use data; income ranges from 1=less than or equal to \$15,000 and 13=greater than \$235,000. We also account for parental occupation, designated as whether students' fathers and/or mothers were employed in STEM fields. Previous research indicates that students' academic performance matters (e.g., Wang et al., 2013). Therefore, we included students' high school mathematics and science course grade point averages (GPA) to measure students' observed ability, in distinction to their perceived abilities. Finally, as indicators of high school preparation for STEM coursework in

college, we also include indicators of highest mathematics and science courses taken. Detailed coding information is provided Table 7 in Appendix.

High school and college characteristics High school characteristics may affect students' STEM plans in college (Legewie & DiPrete, 2014). Therefore, we controlled for high school characteristics by using variables representing high school types and urbanicity, as shown in Table 7. High school types include public, Catholic, and other private high schools. Urbanicity indicates whether the school is an urban, suburban, town, or rural area.

Classroom teachers likely have direct influence on their students' STEM ambitions as well. Aaronson et al. (2007) found math teacher quality was positively associated with students' math scores. Additionally, since academic interest is one of the main independent variables and previous research (Frenzel et al., 2010) indicates interest can be influenced by students' relationships with teachers, we include students' ratings of their experiences with their math/science teachers in our analyses. In the base-year student survey, these two variables were measured by nine items. Similar processes for generating these variables were conducted by testing Cronbach's reliability for students' experiences with their mathematics ($\alpha=0.89$) and science ($\alpha=0.89$) teachers (see details Table 10 in Appendix). Higher values of each variable are associated with students having better experiences with their math/science teacher.

Previous studies have found that postsecondary institutions matter (e.g., Phelps et al., 2018; Wang, 2013b). We included students' first college type to account for postsecondary characteristics. This variable is a dichotomous variable, coded as 1=2-year or less and 0=4-year.

Limitations

Academic interests may be situational, Hidi and Renninger note (2006), affected by external environments. It is possible that during the period of the base-year survey, students were interested in mathematics and science, but then lost interest before entering college, influencing their major choice. Our measures may not fully represent students' true academic interests at the point they were making college major decisions, which could bias the estimated effects of academic interests on STEM major choice. Indeed, research suggests deep engagement and/or passion in a STEM subject could be transformative to sustaining interest (e.g., Eccles & Wigfield, 2020; Schneider et al., 2016). Our assessment of this variable is limited to the effect of self-reported interest at the start of high school, which indeed may be malleable to change in accordance with the variables we measure and control for, including subsequent course taking and school context variables.

¹ As another test of our scales, we report at the end of the results section on a sensitivity analysis, substituting the NCES items for our own scales; this series of models less effectively predicted our dependent variables than the scales we use and report on in the text, further confirming the appropriateness of our design.

Another limitation is the nature of secondary self-reported large-scale data. Our measure of student gender is drawn from a measure of biological sex (male/female) without options beyond this binary during the first three study waves. We note this again later in reporting out the results. In addition, both perceived abilities and academic interests were self-reported by adolescents, which may influence the survey accuracy. Additionally, HSLs: 09 is not explicitly designed for the measure of psychological constructs and therefore may not be able to measure perceived abilities and academic interests fully. However, we argue that the choice of dataset allows us to combine reliable measures (as noted by the scores reported above) and a nationally representative longitudinal cohort, to assess how interests and perceived ability in mathematics and science affect STEM choices over time.

Analytic approaches

We addressed our research questions and hypotheses through logistic regression due to the categorical nature of our dependent variables—binary logistic regression when predicting “STEM” major choice and a multinomial logistic regression with our more detailed categorical dependent variable. Regression equations and corresponding hypotheses are described below. For the first research question, the dependent variable is students’ college STEM choices. The main independent variables were perceived mathematical and scientific abilities and interests, while adjusting for other demographic, high school, and classroom characteristics. Two main hypotheses were proposed for the first research question:

H1a: Students with high perceived mathematical/scientific abilities are more likely to choose a STEM major.

H1b: Students with high interest in mathematics/science are more likely to choose a STEM major.

To test H1a, we started with an initial model that only included students’ STEM major choice as the dependent variable and perceived mathematical/scientific abilities as the independent variable. Then, we added student characteristics, high school/college characteristics, and classroom characteristics one by one to test our hypotheses. For simplification, we present the final model as follows:

$$\text{logit}\left(\frac{P_{STEM}}{1 - P_{STEM}}\right)_i = \beta_0 + \beta_1 PMA_i + \beta_2 PSA_i + \beta_3 STC_i + \beta_4 SC_i + \beta_5 CC_i, \tag{1}$$

where STEM = whether student choose a college major in STEM fields; PMA_i = the perceived mathematical ability of student i ; PSA_i = the perceived scientific ability of student i ; STC_i = a vector of student i ’s demographic

information such as gender, race/ethnicity, parental education, parental occupation, family income, student preparation and observed ability in math and science in high school; SC_i = a vector of student i ’s high school/college information including school type, school’s urbanicity, and college type; and CC_i = a vector of student i ’s classroom experiences with their mathematics/science teachers.

The procedure of testing H1b is similar to the process when examining the effect of perceived abilities on students’ STEM major choice. The only difference is substituting perceived ability variables for interest variables. Thus, our final model for H1b was:

$$\text{logit}\left(\frac{P_{STEM}}{1 - P_{STEM}}\right)_i = \beta_0 + \beta_1 MI_i + \beta_2 SI_i + \beta_3 STC_i + \beta_4 SC_i + \beta_5 CC_i, \tag{2}$$

where MI_i = student i ’s mathematics interest; and SI_i = student i ’s science interest.

The second research question assesses whether perceived abilities or interests could be more predictive of students’ STEM major choice. To address this research question, we added both sets of perceived abilities and interests into one model:

$$\text{logit}\left(\frac{P_{STEM}}{1 - P_{STEM}}\right)_i = \beta_0 + \beta_1 PMA_i + \beta_2 PSA_i + \beta_3 MI_i + \beta_4 SI_i + \beta_5 STC_i + \beta_6 SC_i + \beta_7 CC_i. \tag{3}$$

Though the scale used for perceived abilities and interest is the same, we reported regression coefficients and odds ratios to compare the relative effects of mathematics and science perceived ability and interest variables on the dependent variable. In addition, we reported the model statistic (F -test) to compare overall model strength. We additionally assessed whether perceived abilities and interest were too highly correlated with each other or otherwise linearly related. Specifically, we estimated an ordinary least squares model to check the variance inflation factor (VIF) for multicollinear relationships among our variables, before entering them into the logistic regression.

The third research question seeks to examine whether the relationships between perceived abilities and interests and STEM major choices vary by gender and race/ethnicity. We generated multiple interaction terms² to

² Specific interaction variables between gender/race and perceived abilities/interests were generated. They are: Gender * PMA, Gender * PSA, Gender * MI, Gender * SI, White * PMA, Latino * PMA, Black * PMA, Asian * PMA, Other/Multi * PMA, White * PSA, Latino * PSA, Black * PSA, Asian * PSA, Other/Multi * PSA, White * MI, Latino * MI, Black * MI, Asian * MI, Other/Multi * MI, White * SI, Latino * SI, Black * SI, Asian * SI, Other/Multi * SI.

see whether these intentional relationships vary by gender and race (Kaufman, 2018). We also estimated a model with state fixed effects, controlling for unobserved factors at the state level such as STEM employment opportunities and access to 4-year colleges. This model did not improve upon the prior model’s explanatory power, nor did it change the trajectory of the results. Therefore, to conserve space and focus on key findings, we do not report them here.

To answer our last research question, we estimated multinomial logistic regression models using the same covariates used to answer research questions 2 and 3, but using a three-category rather than a binary dependent variable, where Non-STEM is the reference group, and we assessed how the results vary between the (aggregated) STEM results reported on in Table 4 and STEM major clusters reported on in Table 6: Computer/Engineering Sciences and Social/Behavioral, Natural, and Other Sciences.

In summary, RQ1 is assessed using Eqs. (1) and (2), and RQ2 is evaluated using Eq. (3). RQ3 is subsequently investigated with the use of interaction terms added to Eqs. (1) and (2). RQ4 is examined by changing the analytic approach used for RQ2 and 3 to a multinomial logistic regression rather than a binary logistic regression while retaining the predictors, and changing the dependent variable to the recoded three-category major variable in Eq. (3).

Results

Descriptive results

Table 1 reports the descriptive statistics for control variables for the weighted analytic sample. Approximately 53.6% of sampled students identify as women. White students accounted for 54.6% of the analytic sample, while Latina/o/x, Black, Asian, and Multiple/Other Race were 20.3%, 12.2%, 4.5%, and 8.5%, respectively. Notably, there was less than a one percentage point difference between the share of students’ fathers (15.3%) and mothers (14.5%) working in STEM occupations. While on average most students completed Intermediate and many completed Advanced Math, most students had not completed a year of physics by the time they entered postsecondary school (as the mean of 0.5 is halfway between 0=No Physics and 1=General Physics). Even among students in our analytic sample of postsecondary degree declarers, this is a striking indicator of U.S. students’ under-preparation in Physics, a key gateway course for most postsecondary STEM majors. With respect to school characteristics, 90.3% of students in the sample attended public high schools, and 23.9% of schools were in rural areas. In addition, students reported that they had good experiences with their mathematics and science teachers; the

Table 1 Descriptive statistics of control variables for the analytic sample with multiple imputation

	Mean	S.E	Min	Max	% Missing
Student characteristics					
Gender					
Male	46.4%	0.0	0.0	100.0	0.2%
Female	53.6%	0.0	0.0	100.0	0.2%
Race/ethnicity					
Asian	4.5%	0.0	0.0	100.0	3.8%
Black	12.2%	0.0	0.0	100.0	3.8%
Latina/o/x	20.3%	0.0	0.0	100.0	3.8%
Multiple/Other Race	8.5%	0.0	0.0	100.0	3.8%
White (Reference)	54.6%	0.0	0.0	100.0	3.8%
Family characteristics					
Family income	4.61	0.1	1.0	13.0	19.8%
Father’s occupation (STEM)	15.3%	0.1	0.0	100.0	34.6%
Mother’s occupation (STEM)	14.5%	0.1	0.0	100.0	26.3%
Parental education					
4-year college degree (Reference)	44.1%	0.0	0.0	100.0	19.7%
No 4-year college degree	55.9%	0.0	0.0	100.0	19.7%
Student preparation and performance					
Highest math course taken	1.61	0.0	0.0	2.0	4.4%
Highest science course taken (Physics pipeline)	0.5	0.0	0.0	2.0	4.4%
HS math course GPA	2.6	0.0	0.0	4.0	4.7%
HS science course GPA	2.7	0.0	0.0	4.0	4.9%
High school and college characteristics					
School types					
Public (Reference)	90.3%	0.0	0.0	100.0	0.0%
Catholic	5.1%	0.0	0.0	100.0	0.0%
Other private	4.7%	0.0	0.0	100.0	0.0%
School location					
Urban (Reference Group)	31.9%	0.0	0.0	100.0	0.0%
Suburban	34.7%	0.0	0.0	100.0	0.0%
Town	10.1%	0.0	0.0	100.0	0.0%
Rural	23.2%	0.0	0.0	100.0	0.0%
Type of college first attended					
Four-year (Reference)	63.0	0.9	0.0	100.0	1.0%
Two-year	37.0	0.9	0.0	100.0	1.0%
Classroom characteristics					
Experiences with math teacher	3.1	0.0	1.0	4.0	15.7%
Experiences with science teacher	3.1	0.0	1.0	4.0	20.9%

The analytic sample size is $n = 11,560$. S.E. refers to Standard Error. Survey weights ($w4w1stup1$) are applied to account for students and parents’ nonresponses to enhance the external validity. This table is using percentage (0–100 scale) to describe the dichotomous variables defined in Table 7 in Appendix to ease interpretation. In the analysis, this study still uses the actual scale (0–1). In addition, the means were rounded to the nearest tenth decimal to comply with NCES restricted-data regulations. Missing data figures reported were generated from Stata 16’s “misstable” command prior to imputation

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLSL: 09), Base Year, Student Survey, 2009; Parent Survey, 2009; Second Follow-up, Student Survey, 2016; Common Core of Data 2005–06, Private School Survey, 2005

Table 2 Dependent variable distribution in the analytic sample with multiple imputation, by gender and race ($N = 11,560$)

Research Questions 1–3: Predicting STEM Major Selection							
	Gender $\chi^2 = 438.7$, $p < 0.001$		Race $\chi^2 = 301.2$, $p < 0.001$				
	Boys	Girls	Asian	Black	Latina/o/x	Multiple/Other	White
STEM	1680 32.0%	980 15.5%	500 42.5%	170 16.2%	310 19.5%	230 22.3%	1440 21.6%
Non-STEM	3560 68.0%	5340 84.5%	680 57.5%	890 83.8%	1280 80.5%	820 77.7%	5230 78.4%
Total	5240	6320	1180	1060	1590	1050	6670
Research Question 4: Predicting Computing/Engineering vs. Other Sciences							
	Gender $\chi^2 = 375.9$, $p < 0.001$		Race $\chi^2 = 12.8$, $p < 0.01$				
	Boys	Girls	Asian	Black	Latina/o/x	Multiple/Other	White
Computing/Engineering	1130 67.5%	280 28.6%	250 50.4%	90 51.5%	170 56.5%	100 44.0%	790 55.0%
Social/Behavioral, Natural, and Other Sciences	550 32.5%	700 71.4%	250 49.6%	80 48.6%	140 43.6%	130 56.0%	650 45.0%
Total	1680	980	500	170	310	230	1440

Total Number refers to the total number of observations in each gender and/or race group. The survey weight (w4w1stup1) was applied to account for nonresponse and the stratified sampling design, to enhance the external validity. The N in each subgroup was rounded to the nearest ten and the percentages rounded to the nearest tenth decimal, to comply with NCES restricted-data regulations. The total number of cases in each subgroup may not be exactly equal to that of the analytic sample because of this rounding. Statistical significances within each gender and race group were assessed using Chi-square tests, with the test statistic and p -value reported for each variable

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS: 09), Base Year, Student Survey, 2009; Second Follow-up, Student Survey, 2016

weighted means were both 3.1 on a scale where 4 means “strongly agree” and 1 means “strongly disagree”.

STEM disparities by gender and race

Postsecondary major outcomes

Table 2 reports on the dependent variables for Research Questions 1–4, as they vary by gender and race. We included Chi-square tests to assess significant differences in the distribution of STEM/Non-STEM (RQ1-3) and Computer/Engineering vs. Other Sciences (RQ4) major selection among students. First, with respect to the aggregated STEM variable, we observed significant differences by gender ($\chi^2 = 438.7$, $p < 0.001$) and to a lesser extent by race ($\chi^2 = 301.2$, $p < 0.001$). Specifically, we found that only 15.5% of women chose STEM majors as compared to 32.0% of men. With respect to students from race groups not historically underrepresented in STEM, 21.6% of White students and 42.5% of Asian students chose STEM majors. Among STEM underrepresented race groups, the percentage of Latina/o/x, Black, and Other/Multiracial students choosing STEM were 19.5%, 16.2%, and 22.3%, respectively. Asian students were the only group with

a meaningful—and positive—difference as compared to their peers.

Next, we turned to our review of disparities in Computing/Engineering vs. Other Science major selection, which we found—similar to the pattern above—varies significantly by gender ($\chi^2 = 375.9$, $p < 0.001$) and (less so) by race ($\chi^2 = 12.8$, $p < 0.01$). Here, we filtered out non-STEM students to compare Computing/Engineering majors to those in the Natural Sciences (e.g., Biology, Physics), Social/Behavioral Sciences, and often applied Other Sciences (e.g., Agriculture, Architecture). We found that 2/3 of boys in STEM fields selected Computing/Engineering majors, as compared to just over 1/4 of girls. With respect to race, we found limited differences between groups in who selects Computing/Engineering, such that variation did not exceed a 13-percentage point difference (between Multiple/Other at 44.0% and Latina/o/x at 56.5%). 55.0% of White students, 51.5% of Black students, and 50.4% of Asian students select Computing/Engineering vs. Other STEM fields. Across these analyses, we found that race disparities existed but were less severe than the gender disparities.

Gender and race intersections: Differences in key psychological predictors

To better understand gender and race differences in students' perceived abilities and interests in mathematics and science, we constructed a series of gender–race subgroups such as White male and Black female to conduct the Hedges' g effect size test of mean differences.³ Hedges' g is appropriate when the sizes of subgroups are different (Kemp and colleagues, 2010). Using White male as the reference group, results reported in Table 3 showed that White females had significantly lower perceived mathematical and scientific abilities, which was consistent with the Riegle-Crumb, King, et al. (2019) and Riegle-Crumb, Morton, et al. (2019) findings that males' math and science self-efficacies were statistically significantly higher than that of females. It might be that negative gender stereotypes socialize many girls to believe they were not good at mathematics and science (Martinot & Désert, 2007). Notably, Latino males' mathematics interest and Black males' interest in mathematics and science were significantly higher than that of White males, in line with Riegle-Crumb et al. (2010).

However, Latino and Black males reported lower perceived scientific ability than their White male counterparts. This suggests that racially minoritized students may have strong interest but low self-efficacy in these STEM-related subjects. Similarly, Black females reported lower perceived scientific ability than White males (2.7 vs. 2.9; 0.43 SD difference). However, their mathematical interests were significantly higher than White male peers (3.0 vs. 2.9; 0.26 SD difference). Our descriptive findings—not yet adjusting for other factors—were consistent with previous findings that descriptive gender and racial-ethnic gaps exist in STEM fields, prior to adding controls and covariates as we did in the next steps of our analysis.

Effects of perceived abilities and academic interests on postsecondary STEM major

RQ1: Do perceived abilities and academic interests predict STEM major choice?

To measure the longitudinal relationship and potential interaction between students' postsecondary STEM major selection and their earlier perceived abilities and academic interests, we tested multiple models. Initial

models included only the outcome variable and perceived mathematical and scientific abilities. Next, we added student characteristics, and then high school/college and classroom characteristics. Across all models, perceived mathematical and scientific abilities both positively influenced STEM major choice. For simplicity, Table 4 shows only the full models, which included student, high school/college, and classroom characteristics. Unstandardized coefficients were reported in addition to odds ratios, which indicated the effect of the predictor on the dependent variable (STEM major choice), such that the relative effect on the odds of the event occurring (choosing a STEM major) could be compared across the predictors, irrespective of the units of each predictor variable.

Beginning with Model 1, we found a one unit increase in students' perceived mathematical ability scale was associated with a 62% increase in the odds of choosing STEM disciplines, holding other variables constant. Similarly, a one unit increase in students' perceived scientific ability scale was associated with an 61% increase in the odds of majoring in STEM fields. The odds of choosing STEM majors were 0.59 times lower for HSLs cohort girls than for otherwise similar boys. Asian students were 1.49 times more likely than White students to major in STEM fields.

Students' high school physics course completion could significantly and positively predict STEM major choice, with $OR=1.72$, $p<0.001$. A significantly negative relationship was found between private schools and STEM degrees. Our results showed that the odds of choosing STEM majors were lower for students from private schools than those of students from public schools ($OR=0.61$, $p<0.001$).

Next, we examined the effects of mathematics and science interest in high school on STEM major choice. The results of Model 2 in Table 4 show that both mathematics and science interests significantly and positively predicted students' STEM choices, with $OR=1.23$, $p<0.05$ and $OR=1.30$, $p<0.05$, respectively. A one unit increase in mathematics interest was associated with a 23% increase in the odds of majoring in STEM, all else constant. Gender remained significant and even more negative in this model, as girls see a decrease of 64% in the odds of choosing STEM majors ($p<0.001$). A one unit increase in science interest was associated with a 30% increase in the odds of choosing STEM majors. In addition, high school physics course completion was positively associated with STEM major choice ($OR=1.78$, $p<0.001$) as is, to a lesser degree, high school science GPA ($OR=1.34$, $p<0.05$). Again, attending a private school was (even more) negatively associated with STEM major choice ($OR=0.51$, $p<0.001$).

³ In discussing these analyses, we use “female” and “male” as terms because of the focused attention on students' experiences from adolescence into young adulthood. Students' ability and interest ratings are measured in high school and are assessed in relation to their entry to STEM majors in postsecondary school, when they are adults. We recognize the limitations of female/male as terms that have been associated with binary notions of biological sex, in distinction to gender, and intentionally use girl/woman and boy/man in the majority of the manuscript.

Table 3 Descriptive statistics for students' math/science-related characteristics: mean, standard deviations, and effect sizes

Variables	White male N = 3040	White female N = 3600	Latina/o/x male N = 650	Latina/o/x female N = 860	Black male N = 430	Black female N = 560	Asian male N = 510	Asian female N = 580	Multi/Other male N = 420	Multi/Other female N = 550
Perceived ability										
Perceived Math Ability (Effect size)	3.0 (.57)	2.9* (.62)	3.0 (.56)	2.8* (.61)	3.0 (.59)	3.0* (.59)	3.3* (.47)	3.1 (.60)	3.0 (.64)	2.9* (.59)
Perceived Sci ability (Effect size)	2.9 (.55)	2.8* (.56)	2.8* (.56)	2.7* (.52)	2.8* (.55)	2.7* (.53)	3.0* (.50)	2.9 (.61)	2.9* (.63)	2.9 (.56)
Academic interests										
Math interests (Effect size)	2.9 (.68)	2.9 (.64)	3.0* (.61)	2.9 (.68)	2.9* (.66)	3.0* (.67)	3.1* (.64)	3.0* (.65)	2.8 (.71)	2.7* (.85)
Science interests (Effect size)	2.9 (.73)	2.9 (.68)	2.9 (.67)	2.9 (.68)	3.0* (.69)	2.8 (.72)	3.1* (.65)	3.1* (.69)	2.8 (.82)	2.7* (.83)
Experiences with teachers										
Math class teacher (Effect size)	3.1 (.54)	3.1 (.54)	3.2 (.51)	3.1 (.55)	3.1 (.65)	3.3* (.48)	3.3* (.49)	3.2* (.47)	3.0* (.57)	3.2 (.50)
Science class teacher (Effect size)	3.1 (.57)	3.1 (.54)	3.1 (.62)	3.1* (.55)	3.1 (.55)	3.0* (.60)	3.3* (.60)	3.2* (.48)	3.0* (.60)	3.0* (.58)

Standard deviations are in parentheses. Effect sizes were calculated as mean differences between two groups, using Hedges' *g* (reference = White male). An asterisk means the given mean is statistically significantly different at the 0.05 level. The survey weight (w4w1stup1) was applied to account for nonresponse, to enhance external validity, and multiply imputed data are used. To comply with NCES restricted-data regulations, means were rounded to the nearest tenth decimal, and sample Ns were rounded to the nearest ten. When comparing subgroup means, some significantly different means may appear identical because of rounding

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS:09), Base Year, Student Survey, 2009

Table 4 Logistic regression results for perceived abilities and academic interests on STEM major choice

Dependent variable	Perceived abilities only (1)		Academic interests only (2)		Perceived abilities and academic interests (3)	
	<i>b</i>	OR	<i>b</i>	OR	<i>b</i>	OR
STEM degree attainment						
Perceived abilities						
Perceived mathematical ability	0.48*** (0.11)	1.62			0.46*** (0.12)	1.58
Perceived scientific ability	0.47*** (0.10)	1.61			0.43*** (0.10)	1.53
Academic interests						
Math interest			0.21*(0.08)	1.23	0.07 (0.10)	1.07
Science interest			0.29* (0.11)	1.30	0.12 (0.12)	1.12
Student characteristics						
Gender (Reference = Boys)						
Girls	− 0.89*** (0.10)	0.41	− 1.00*** (0.09)	0.36	− 0.90*** (0.10)	0.041
Race/ethnicity (Reference = White)						
Asian	0.58** (0.18)	1.49	0.39* (0.19)	1.47	0.39* (0.18)	1.48
Black	0.06 (0.15)	1.04	0.03 (0.17)	1.03	0.03 (0.16)	1.03
Latina/o/x	0.22 (0.12)	1.24	0.20 (0.13)	1.22	0.20 (0.13)	1.23
Multiple and other race	0.08 (0.14)	1.22	0.18 (0.14)	1.19	0.12 (0.15)	1.13
Family characteristics						
Family income	− 0.01 (0.02)	0.99	− 0.00 (0.02)	1.00	− 0.01 (0.15)	0.99
Parental college degree	0.10 (0.10)	1.10	0.12 (0.09)	1.12	0.07 (0.09)	1.11
Father's occupation (STEM)	0.16 (0.09)	1.18	0.20* (0.09)	1.22	0.16 (0.09)	1.18
Mother's occupation (STEM)	0.10 (0.10)	1.11	0.13 (0.10)	1.14	0.11 (0.10)	1.11
Student preparation and observed ability in high school						
Highest HS mathematics course	0.16 (0.11)	1.17	0.17 (0.11)	1.19	0.16 (0.11)	1.17
Highest HS science course	0.54*** (0.07)	1.72	0.57*** (0.07)	1.78	0.54*** (0.07)	1.71
HS math course GPA	0.06 (0.12)	1.07	0.14 (0.10)	1.15	0.07 (0.12)	1.07
HS science course GPA	0.23 (0.13)	1.26	0.30* (0.13)	1.34	0.23 (0.13)	1.26
School characteristics						
High school type (Reference = Public)						
Catholic	− 0.13 (0.11)	0.87	− 0.19 (0.11)	0.82	− 0.14 (0.11)	0.87
Other private	− 0.61*** (0.12)	0.54	− 0.68*** (0.12)	0.51	− 0.62*** (0.12)	0.54
College type first attended						
Two-year college	− 0.16 (0.11)	0.85	− 0.17 (0.11)	0.85	− 0.16 (0.11)	0.85
High school location (Reference = Urban)						
Suburban	− 0.04 (0.10)	0.96	− 0.04 (0.10)	0.96	− 0.04 (0.10)	0.96
Town	0.08 (0.15)	1.09	0.05 (0.15)	1.05	0.08 (0.15)	1.09
Rural	− 0.17 (0.11)	0.85	− 0.20 (0.11)	0.82	− 0.17 (0.11)	0.85
Classroom characteristics						
Math teacher	− 0.04 (0.09)	0.96	− 0.01 (0.10)	0.99	− 0.09 (0.11)	0.93
Science teacher	− 0.02 (0.11)	0.97	− 0.15 (0.10)	0.94	− 0.10 (0.11)	0.91
Constant	− 0.01***		− 3.91***		− 6.02***	
F test	26.11***		19.09***		24.26***	
Sample N	11,560		11,560		11,560	
Population N	2,713,260		2,713,260		2,713,260	

p* < 0.05, *p* < 0.01, ****p* < 0.001. Standard errors are shown in parentheses. OR represents odds ratios, where numbers greater than zero represent a positive increase in the odds an event will occur. *N* was rounded to the nearest ten to comply with NCES restricted-data regulations

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS: 09), Base Year, Student Survey, 2009; Parent Survey, 2009; 2013 Update, Second Follow-up, Student Survey, 2016; Common Core of Data, Private School Survey, 2005

Table 5 Correlations and multicollinearity between perceived abilities and interests

Variable names	1	2	3	4	VIF	Tolerance
1. Perceived mathematical ability	1.00				1.53	0.65
2. Perceived scientific ability	0.37	1.00			1.53	0.65
3. Math interest	0.51	0.19	1.00		1.42	0.70
4. Science interest	0.18	0.51	0.27	1.00	1.42	0.70

N = 11,560. Variance inflation factor (VIF) was obtained from OLS regression to test for potential multicollinearity. *N* was rounded to the nearest ten, to comply with NCES restricted-data regulations

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL: 09), Base Year, Student Survey, 2009

RQ2: Comparison testing: Perceived abilities or academic interests?

Research Question 2 examined whether perceived abilities or interests more strongly predict students’ STEM major choice, associated with the results of Table 4’s Model 3. Multiple tests yielded the same answer. Looking across our three models—perceived abilities only (1), interests only (2), and abilities + interests (3)—the *F*-tests all showed a significant predictive relationship on STEM major choice. However, the perceived abilities only model was the strongest (*F* = 26.1), and the interests only model was the weakest (*F* = 19.1). Next, the independent variables’ *t*-test results also favored perceived abilities over interests as predictors of STEM major choice. When both perceived abilities and interests were examined together in Model 3, only perceived mathematical/scientific abilities remained significant, positively predicting STEM degree major fields. Adding mathematics and science interests to the model only slightly diluted the predictive power of perceived mathematics and scientific ability in Model 3 as compared to the simpler Model 1. In both cases, the ability measures were significant beyond the *p* < 0.001 level, with changes being less than 0.05 in magnitude for the slope coefficient and less than 0.10 for odds ratios. In summary, the limitations of these measures notwithstanding, we found that perceived abilities mattered more than interests as predictors of students’ STEM degrees, all else equal.

Before closing our discussion of the independent effects of abilities and interests on STEM major choice, we added an additional analysis to confirm this finding and assessed the degree to which perceived abilities and interests might be highly correlated or non-linear in their relationship. The correlations between perceived mathematical/scientific abilities and math/science interests were both 0.51, moderate correlations as suggested by Evans (1996). Next, we estimated an OLS regression model to obtain the variance inflation factors (VIF), to rule out any potential multicollinearity. As seen in Table 5, all the VIFs were less than 3, therefore multicollinearity was not a concern (Thompson et al., 2017).

Having ruled out such concerns, we could then infer with confidence that perceived abilities were stronger and more important predictors of students’ STEM majors than academic interests. This finding recalled our earlier descriptive finding, that there was more significant race–gender variation on perceived than academic interest (Table 3). Altogether, we found that perceived abilities played a more crucial role than academic interests when students chose their college majors.

RQ3: Do these relationships vary by gender and race?

To examine whether these relationships between perceived abilities, interests, and STEM major choice varied by gender and race, we added interaction terms into the respective models. All other variables’ predictive relationships with STEM degree fields remained consistent with Model 1 and Model 2 as reported earlier in Table 4. After estimating these two models with gender and race interaction terms, significant interaction terms were found for gender only. Specifically, we found a positive main effect for male (*b* = 3.19, *p* < 0.001) and perceived science ability (*b* = 0.71, *p* < 0.001), and a negative interaction term (*b* = − 0.04, *p* < 0.05). This indicated that the nature of these relationships varied to a degree by gender (favoring boys) but not by race. We investigated further gender and race interactions in response to research question #4.

RQ4: Assessing Computing/Engineering major choice vs. other STEM fields

Given the heterogeneity of disciplinary fields represented within “STEM”, we examined how sensitive our earlier findings were to an alternative specification, which examined a three-level dependent variable. Multiple logistic regression was used to assess the effects of mathematical/scientific perceived abilities and interests on the choices of postsecondary degrees in Computing/Engineering fields, Other STEM fields, and non-STEM fields (the reference group). We reported the findings from the full model in Table 6, using Model 3 from Table 4 (abilities + interests) to investigate differences between the predictive patterns.

The primary findings held in this more nuanced analysis. Academic interest did not reach significance for either STEM major cluster. Perceived ability remained highly significant as a predictor of students' attainment of Computing/Engineering degrees ($RRR_{\text{math ability}} = 1.63$, $p < 0.01$; $RRR_{\text{science ability}} = 1.50$, $p < 0.001$) and other STEM fields ($RRR_{\text{science ability}} = 1.50$, $p < 0.001$). We could see how the gender disparity in STEM varied widely and indeed flipped depending on which STEM cluster we examined, as girls' probability of majoring in Computing/Engineering was strongly negative ($RRR = 0.25$, $p < 0.001$) but their probability majoring in other STEM fields that were generally less applied and less sex-segregated was highly positive, all else equal ($RRR = 2.35$, $p < 0.001$). Race matters differently in each, and only in part. Identifying as Latina/o/x as compared to white was positively associated with Computing/Engineering major choice ($RRR = 1.48$, $p < 0.05$), and identifying as Asian was positively associated with STEM major choices in the comparatively less applied Other STEM category ($RRR = 1.65$, $p < 0.001$). Private schools remained a significant and negative predictor, but only with Computing/Engineering majors ($RRR = 0.68$, $p < 0.01$). High school science GPA was predictive only of the non-Computing/Engineering STEM cluster ($RRR = 1.25$, $p < 0.01$), and course taking was insignificant across categories.

Returning to Research Question #3, we reported on interaction models with figures for clearer demonstration of the predictive and in some cases intersecting relationships between demographic characteristics, abilities, and interests, across these major clusters. More specifically, we used the postestimation margins command in Stata 16 to estimate from the interaction model students' predicted probabilities of choosing a STEM major. Figures 2 and 3 show whether the relationships between perceived abilities/interests and STEM major choice varied for men and women. The slopes did not vary by mathematics/science interest but do vary in perceived mathematical and scientific abilities. In other words, the relationship between mathematics/science interests and STEM major choice (specifically, Computing/Engineering and Other Sciences) did not differ for boys and girls. However, perceived mathematical and scientific ability significantly moderated the relationship between gender and the choice of Computing/Engineering (vs. Non-STEM) degrees, whereby higher perceived ability had a greater effect on boys than girls ($p < 0.05$ for each). The interaction with Social/Behavioral, Natural, and Other STEM fields (vs. non-STEM) was significant in the other direction, where perceived mathematics ability had a more positive effect on girls than boys ($p < 0.05$).

Figures 4 and 5 in Appendix show predicted probabilities by race, by degree field cluster. These lines

consistently had similar trends and slopes. These figures supported our null finding: the predictive relationship between perceived abilities and interests on STEM major choice was stable across students' self-identified race categories.

We added one additional robustness check on our findings, adding a final full re-analysis of our models with the NCES-generated scales for perceived abilities and interests, to assess the merits of our predictors as compared to those made freely available to users of the HSLs dataset (mathematics and science self-efficacy, respectively: $x1mtheff$, $x1scieff$; mathematics and science interest, respectively: $x1mthint$, $x1sciint$). We did so to add further transparency for any studies that might employ these measures for replication or cross-cohort comparisons. The core findings aligned, but our predictor- and model-level statistics were stronger than those using the NCES constructs. For instance, Table 4's full model with perceived abilities and interests had an F -test of 24.3 as compared to its counterpart in the NCES-variable analysis, with an F -test of 19.3. Similarly, Table 6's F -test of 19.5 was stronger than the comparative model which had the same N but a probability of 17.3. Accordingly, we did not report these findings here given the already extensive series of tables and figures included in the manuscript and its supplement. These checks did enhance our confidence in this study's design.

In summary, across our analyses, we found that perceived abilities and interests could significantly and positively predict students' STEM major choice. Notably, perceived mathematical and scientific abilities seemed to be a better predictor than mathematics and science interests. Finally, we found that these relationships did not vary by race. However, the relationship between perceived mathematics and science ability and STEM major choice did vary by gender, and with distinct effects depending on how STEM was defined and categorized.

Discussion

Given persistent disparities in STEM postsecondary majors and degrees, it is imperative to identify malleable factors that may welcome rather than deter students from majoring in these key fields. Our descriptive analyses of U.S. students reported in Tables 1 and 2 show a clear gender gap and more nuanced racial disparities in STEM major choice. The effect size analysis suggests that while Black and Latino male students reported higher mathematics and/or science interest as compared to White male peers, women and some racially minoritized groups still represent a small share of STEM majors. Further exploration of the data reveals that besides perceived abilities and academic interests, students' observed mathematics and science abilities (grades) and father's STEM

Table 6 Predictive relationships between perceived abilities, interests, and STEM major clusters

Dependent variable	Computing and engineering		Social, behavioral, natural, and other sciences		Non-STEM (Reference)	
	RRR	SE	RRR	SE	RRR	SE
Dependent variable						
STEM degree attainment (Specific)						
Perceived abilities						
Perceived mathematical ability	1.63**	0.31	1.08	0.09		
Perceived scientific ability	1.50***	0.19	1.50***	0.13		
Academic interests						
Math interest	1.22	0.17	1.15^	0.08		
Science interest	0.94	0.13	0.99	0.09		
Student characteristics						
Gender (Reference = Boys)						
Girls	0.25***	0.04	2.35***	0.23		
Race/ethnicity (Reference = White)						
Asian	1.24	0.27	1.65***	0.25		
Black	1.00	0.21	1.03	0.14		
Latina/o/x	1.48*	0.26	1.37	0.20		
Multiple and other race	0.83	0.15	1.20	0.14		
Family characteristics						
Family income	0.97^	0.02	0.99	0.01		
Parental college degree	1.12	0.14	0.87^	0.07		
Father's occupation (STEM)	1.12	0.15	0.94	0.12		
Mother's occupation (STEM)	1.19	0.15	1.14	0.11		
Student preparation and observed ability in H.S.						
Highest H.S. math course	1.07	0.13	1.19^	0.11		
Highest H.S. science course	2.02	0.17	1.08	0.09		
H.S. math course GPA	0.98	0.16	0.94	0.07		
H.S. science course GPA	1.23	0.21	1.25**	0.10		
School characteristics						
High school type (Reference = Public)						
Catholic	0.78	0.13	1.11	0.13		
Other private	0.68**	0.10	0.86	0.10		
College type first attended						
Two-year college	1.08	0.15	0.92	0.11		
High school location (Reference = Urban)						
Suburban	0.98	0.13	1.00	0.12		
Town	1.10	0.22	1.28	0.22		
Rural	0.80	0.13	1.20	0.17		
Classroom characteristics						
Math teacher	0.82	0.11	0.85	0.08		
Science teacher	0.96	0.15	1.04	0.11		
Constant	0.01***		0.05***			
F test	19.45***					
Sample N	11,560					
Population N	2,713,260					

^ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors are shown in parentheses. Multinomial logistic regression results are reported. RRR represents relative risk ratios, where numbers greater than zero represent a positive increase in the relative probability of an event occurring. *N* was rounded to the nearest ten to comply with NCES restricted-data regulations

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS: 09), Base Year, Student Survey, 2009; Parent Survey, 2009; 2013 Update, Second Follow-up, Student Survey, 2016; Common Core of Data, Private School Survey, 2005

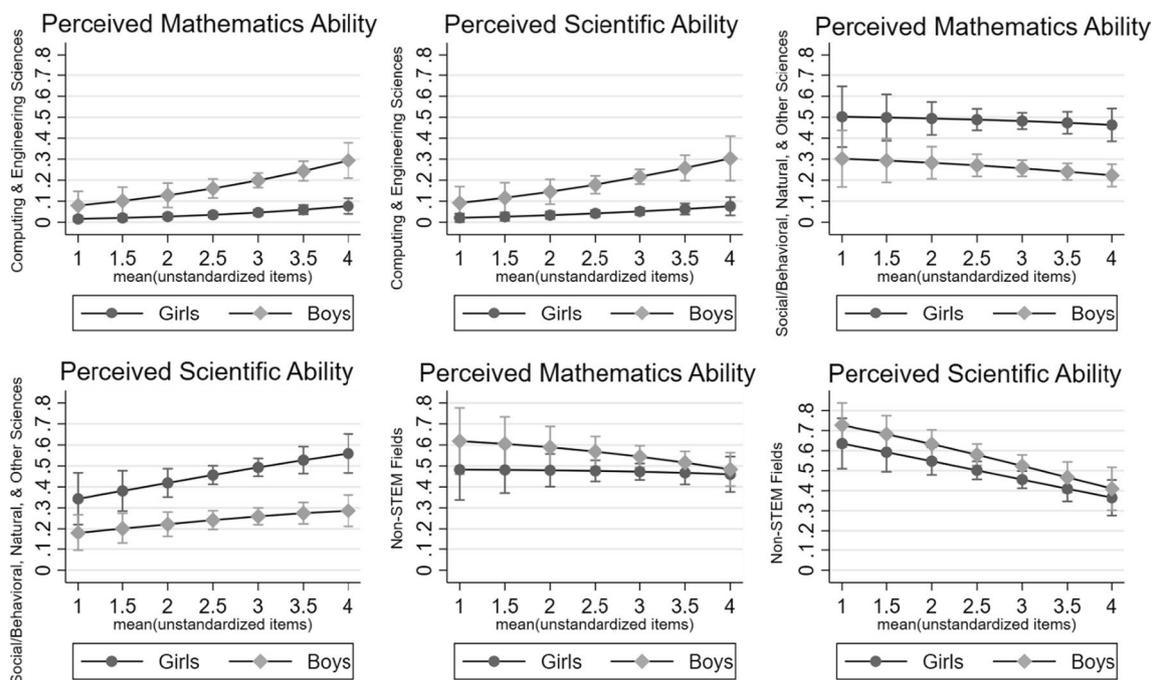


Fig. 2 The relationships between perceived ability and STEM major choice, by gender. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL: 2009–16), Base Year, Student Survey, 2009; Parent Survey, 2009; 2013 Update, Second Follow-up, Student Survey, 2016; Common Core of Data, Private School Survey, 2005

occupation also predict their STEM major choices. Two policy recommendations emerge from these findings.

First, we find that the gender gap in STEM fields remains severe. The results reported in Table 3 indicate that every female subgroup had lower perceived mathematical and scientific abilities than their corresponding same-race male peers (e.g., Black girls had lower perceived abilities than Black boys). This is in line with extant research showing girls’ self-ratings are lower even when test scores are identical to those of boys (Perez-Felkner, et al., 2017). Traditional social norms such as “girls are not as good as boys at math” may continue to influence girls’ cognitive development (Harro, 2000). In our analytic model results, girls are less than half as likely as boys to major in STEM, 3/4 less likely to major in Computing/Engineering fields, and more than twice as likely to major in other STEM fields.

Notably, the moderating effect we observe with our significant interaction terms—where perceived mathematics and scientific ability change the nature of the relationship between gender and major choice—we see the status quo *enhanced*, not lessened. More positive perceptions of one’s ability promotes STEM, Computer/Engineering, and Other STEM major choice for both girls and boys, as we see in Tables 4, 6, and Fig. 2. However, boys see the greatest gains in STEM and Computer/Engineering major choice (as compared to non-STEM fields), even though

they are already overrepresented in these fields. Girls experience greater gains than boys in the relationship between perceived scientific ability and Social/Behavioral, Natural, and Other STEM fields (see Fig. 2). But while these patterns support individual students, at the aggregate level, they reinforce rather than undo existing gender distinctions. It may be that gendered (and racialized) social norms mute the potential effect of academic interest and perceived abilities for students who do not regularly encounter reinforcing motivational forces in their secondary and postsecondary schooling environments.

Socializing contexts in schools, families, and broader culture could be sites to undo the status quo, to allow girls to believe (as boys seem to) that they are positioned to be successful in mathematics and science classrooms and careers that employ these skills. This may be especially important in public schools who enroll most U.S. students, as shown in our descriptive results. Our findings indicate the chances of majoring in STEM are considerably stronger in public high schools, even after controlling for state fixed effects as we did in an earlier analysis. Public high schools appear especially poised for potential interventions to increase STEM interest, which may also foster higher perceived mathematics and science ability in high school and subsequent postsecondary STEM majors. Given the positive association between higher levels of (and indeed any) high school

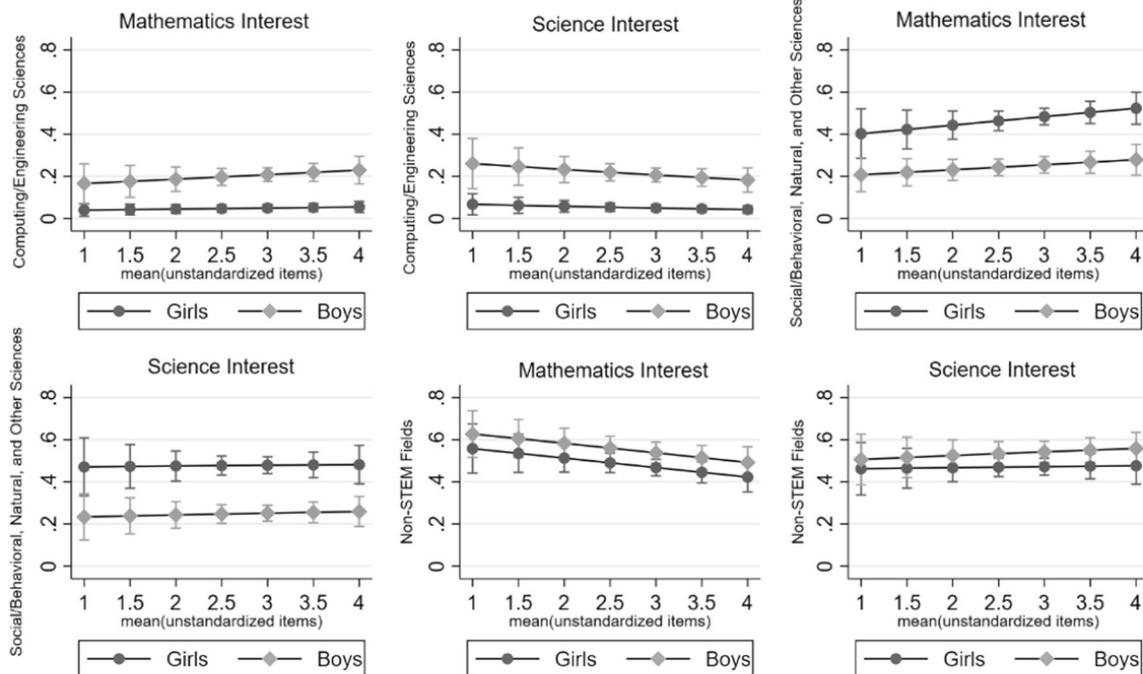


Fig. 3 The relationships between academic interest and STEM major choice, by gender. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL: 2009–16), Base Year, Student Survey, 2009; Parent Survey, 2009; 2013 Update, Second Follow-up, Student Survey, 2016; Common Core of Data, Private School Survey, 2005

physics course taking, making these foundational courses available to students—irrespective of demographic background—appears essential to allow them the opportunity to enter and complete STEM majors across disciplines. Studies of 2-year colleges also suggest both opportunities and challenges for STEM equity in this sector, as gender gaps remain in mathematics-intensive STEM fields (Perez-Felkner et al., 2019) and research attending to gender and race together find additional challenges for racially minoritized women transferring from 2-year to 4-year college STEM programs (Allen et al., 2022).

Our second policy recommendation focuses on financial support for STEM-aspiring students. With respect to socioeconomic challenges that disproportionately affect students of color, St. John and Asker (2003) illustrated that many students with academic preparation cannot enter college because of financial constraints. Quadlin (2017) found that students without financial burdens were more likely to choose liberal arts majors in college, while low-income students may be motivated to pursue college majors associated with higher paying careers, including STEM. However, STEM majors may require additional time and coursework for students whose high schools did not offer as many opportunities, which may present a financial burden. Public and institutional policy interventions can address such challenges by enhancing access to financial aid for

underrepresented students to pursue STEM careers, with benefit for larger economic opportunity.

Conclusion

Using the most recent nationally representative dataset, this study examines the longitudinal effects of self-assessed perceived ability and academic interest in mathematical and scientific domains on high students’ postsecondary STEM major choice. Major findings include that (1) perceived mathematical and scientific abilities positively predict students’ STEM major choice; (2) higher mathematics and science interests increase the likelihood of majoring in STEM fields, but less so than perceived mathematical and scientific abilities, and (3) these relationships vary by gender but do not consistently vary by race/ethnicity. Together, these findings suggest that high school educators ought to focus more on encouraging students in mathematics and science courses and activities, to enhance their perceived mathematical and scientific abilities, ultimately positioning them to potentially enter STEM fields. Additionally, public schools organized to support the ambitions of STEM students of all backgrounds may be better positioned to reduce postsecondary disparities in STEM major choice.

Appendix

See Figs. 4, 5 and Tables 7, 8, 9, 10.

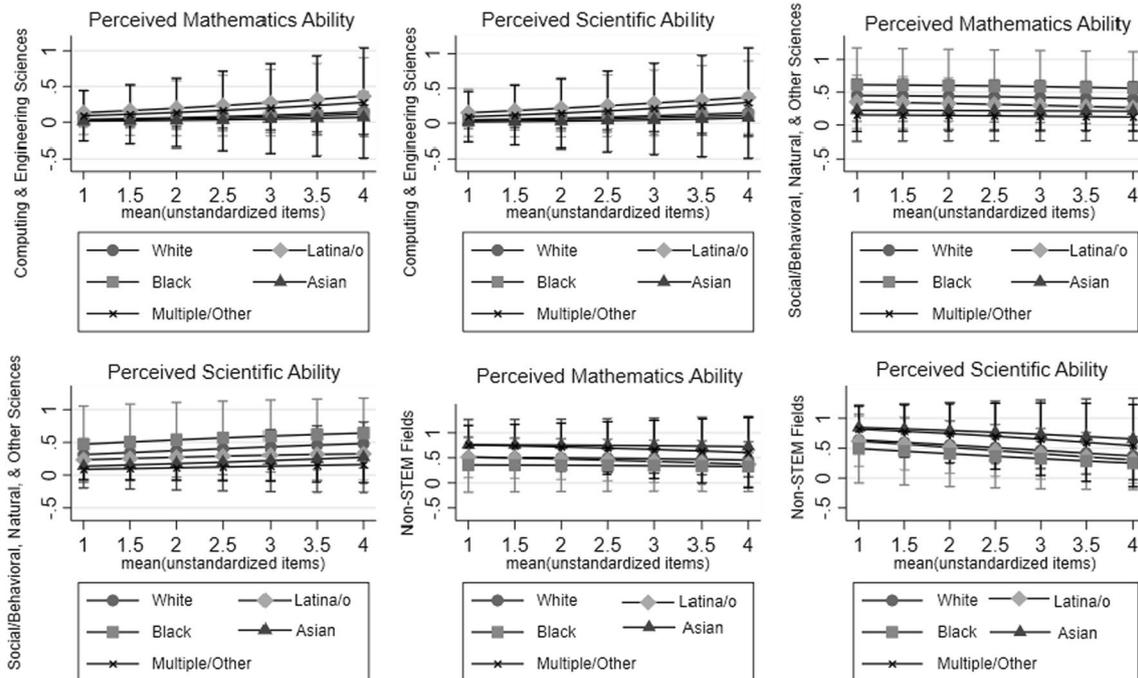


Fig. 4 Predicted probability of selecting specific STEM Majors, by perceived ability and race. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS: 2009–2016), Base Year, Student Survey, 2009; Parent Survey, 2009; 2013 Update, Second Follow-up, Student Survey, 2016; Common Core of Data, Private School Survey, 2005

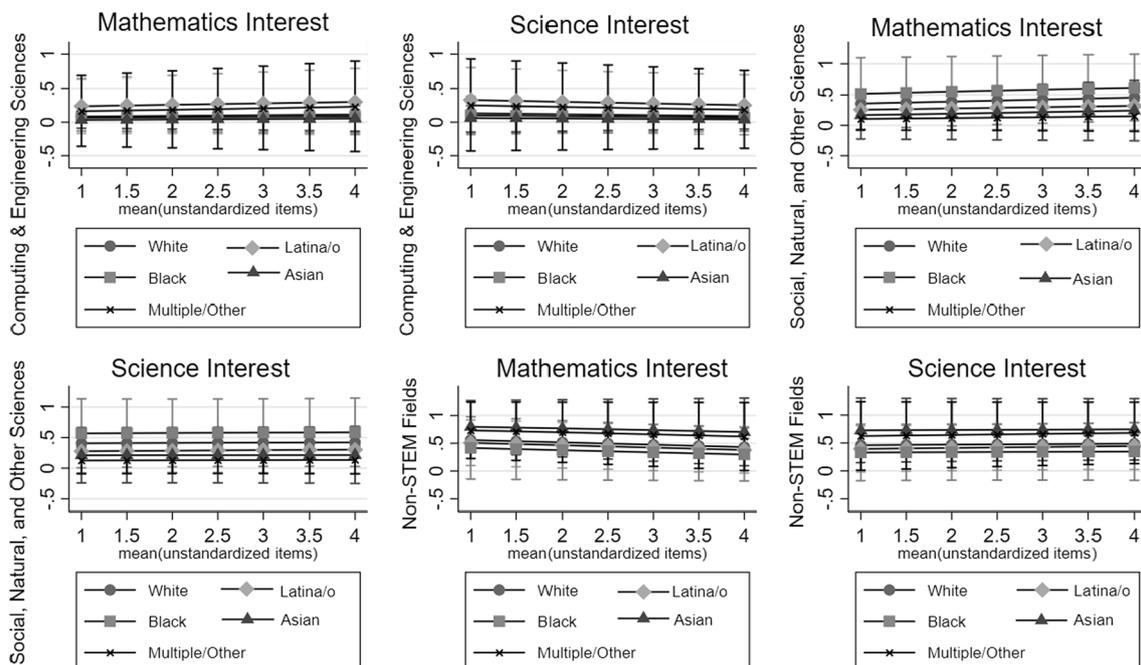


Fig. 5 Predicted probability of selecting specific STEM Majors, by academic interest and race. Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS: 2009–2016), Base Year, Student Survey, 2009; Parent Survey, 2009; 2013 Update, Second Follow-up, Student Survey, 2016; Common Core of Data, Private School Survey, 2005

Table 7 Descriptions of variables: coding, source, type, original name

Variable	Variable description	Source	Variable type	Original variable name
Dependent variable				
STEM major	Whether a student's undergraduate degree field is a STEM major or not (1 = Yes, 0 = No)	HLSL2009 Second Follow-up Student Survey	Dichotomous variable	X4RFDGMJSTEM
STEM major cluster	X4 Reference degree major—all and baccalaureate categories (1 = Non-STEM: General studies and other, Humanities, Business, Education, and Other applied; 2 = Computer/Engineering Sciences; Computer and information sciences and/or Engineering and engineering technology; 3 = Social/Behavioral, Natural, and Other sciences; Biological and physical Sciences, Science technologies, Mathematics, Agriculture, Social Sciences, and Health care fields)	HLSL2009 Second Follow-up Student Survey	Categorical	X4RFDGMJ14Y, which like X4RFDGMJSTEM above uses "X4REFDEG" reference degree identifier as the source, indicating the undergraduate degree or certificate the respondent was working on when last enrolled, as of February 2016
Independent variables				
Gender				
Male	Whether a student's biological sex is male (1 = Yes, 0 = No)	HLSL2009 Base Year Student Survey	Dichotomous variable	X1SEX
Race/ethnicity				
Race	Whether a student identifies as White (1; reference), Asian (2), Black/African American (3), Latina/o/x (4), or Other identity, including multiple non-Hispanic groups (5)	HLSL2009 Base Year Student Survey	Categorical variable	X1RACE
Family characteristics				
Parental education	Whether at least one of a student's parents has a 4-year college degree (1 = Yes, 0 = No)	HLSL2009 Base Year Parents Survey	Dichotomous variable	X1PAREDU
Father's occupation	Whether a student's father's job is a STEM job (1 = Yes, 0 = No)		Dichotomous variable	X1DADDOCC_STEM1
Mother's occupation	Whether a student's mother's job is a STEM job (1 = Yes, 0 = No)		Dichotomous Variable	X1MOMDOCC_STEM1
Family income	X1 Total family income from all sources 2008 (1 = Family income less than or equal to \$15,000; 13 = Family income > \$235,000, with \$20,000 range increments between 1 and 13)		Categorical variable	X1FAMINCOME

Table 7 (continued)

Variable	Variable description	Source	Variable type	Original variable name
School characteristics				
Public (Reference Group)	Whether a student's high school is a public school (1 = Yes, 0 = No)	Common Core of Data 2005–06/Private School Survey 2005–06	Dichotomous variable	X1CONTROL
Catholic	Whether a student's high school is a catholic school (1 = Yes, 0 = No)		Dichotomous variable	
Other Private	Whether a student's high school is another private school (1 = Yes, 0 = No)		Dichotomous variable	
Urbanicity	Whether a student's high school is in a city (1; reference), suburb (2), town (3), or rural area (4)		Categorical variable	X1LOCALE
College characteristics				
Two-year college	College Type: 1st Attended where 0 = 4-year and 1 = 2-year or less	2nd Follow-up, Student Survey, 2016	Dichotomous variable	X4P51SECTOR
Math-related characteristics				
Perceived math ability	Student's perceived math ability, which is measured by seven items using Cronbach's alpha	HLS2009 Base Year/2013 Update Student Survey	Continuous variable	S1MPERSON1, S1MPERSON2, S1MUNDERST, S1MTESTS, S1MTEXTBOOK, S1MSKILLS, S1MASSEXCL
Math observed ability (GPA)	Students' high school math course GPA		Continuous variable	X3TGPAMAT
Mathematics course completion (HS)	X3 Highest level mathematics course taken/pipeline (0 = Algebra 1 or lower; 1 = Geometry, Algebra II, or Trigonometry; 2 = Advanced Math including Calculus)		Ordinal variable	X3THIMATH
Math interests	Student's math interests, which is measured by three items using Cronbach's alpha		Continuous variable	S1MENJOYING, S1MWASTE, S1MBORING
Math teacher	How a student feels about their math teacher, which is measured by nine items using Cronbach's alpha		Continuous variable	S1MTCVALUES, S1MTCHRESPCT, S1MTCFAIR, S1MTCCHCONF, S1MTCCHMISTKE, S1MTCHTREAT, S1MTCCHINTRST, S1MTCCHMFDIFF, S1MTCHEASY

Table 7 (continued)

Variable	Variable description	Source	Variable type	Original variable name
Science-related characteristics				
Perceived science ability	Student's perceived science ability, measured by seven items using Cronbach's alpha	HSL2009 Base Year/2013 Update Student Survey	Continuous variable	S1SPERSON1, S1SPERSON2, S1SUNDERST, S1STESTS, S1STEXTBOOK, S1SSKILLS, S1SASSEXCL
Science observed ability (GPA)	Students' science school course GPA		Continuous variable	X3TGPASCI
Physics course completion (HS)	X3 Highest level physics course taken/ pipeline (0 = No Physics, 1 = General Physics, 2 = Specialty, Advanced, or AP/ IB Physics)		Ordinal variable	X3THIPHY
Science interests	Student's science interests, measured by three items using Cronbach's alpha		Continuous variable	S1SENOYING, S1SWASTE, S1SBORING
Science teacher	How a student feels about their science teacher, which is measured by 9 items using Cronbach's alpha		Continuous variable	S1STCHVALUES, S1STCHRESPCT, S1STCHFAIR, S1STCHCONF, S1STCHMISTKE, S1STCHTREAT, S1STCHINTRST, S1STCHMFDIFF, S1STCHEASY

Source U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HLS: 09), Base Year, Student Survey, 2009; Parent Survey, 2009; 2013 Update; Second Follow-up, Student Survey, 2016; Common Core of Data 2005–06, Private School Survey, 2005

Table 8 Reliability for perceived mathematical/scientific abilities

Items	Strongly disagree			Strongly agree
1. You see yourself as a math/science person	1	2	3	4
2. Others see you as a math/science person	1	2	3	4
3. You really understand math/science assignment	1	2	3	4
4. You are confident to do an excellent job on math/science test	1	2	3	4
5. You are certain that you can understand the most difficult material presented in the textbook in the math/science course	1	2	3	4
6. You are certain that you can master the skills being taught in the math/science course	1	2	3	4
7. You are confident that you can do an excellent job on assignments in the math/science course	1	2	3	4
Reliability	Cronbach's Alpha (α)			
	Perceived Math Ability		Perceived Science Ability	
	0.89		0.87	

Item 3 was negatively worded, all other items were coded as 1 = strongly agree, 4 = strongly disagree. For interpretability, we recoded item 3 as 1 = strongly disagree, 4 = strongly agree. To more precisely generate the Cronbach's reliability measure, we included the full dataset rather than just the analytic sample. Due to item-level missing data, 18,580 and 17,380 observations were tested for perceived mathematical and scientific abilities, respectively

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL: 09), Base Year, Student Survey, 2009

Table 9 Reliability for math and science interests

Items	Strongly disagree			Strongly Agree
1. You are enjoying this math/science class very much	1	2	3	4
2. You think this math/science class is a waste of your time	1	2	3	4
3. You think this math/science class is boring	1	2	3	4
Reliability	Cronbach's Alpha (α)			
	Math Interests		Science Interests	
	0.78		0.81	

Items 2 and 3 were negatively worded. All other items were coded as 1 = strongly agree, 4 = strongly disagree. For interpretability, we recoded items 1 and 2 as 1 = strongly disagree, 4 = strongly agree. Due to the missing data, 18,960 and 17,440 observations are tested for math and science interests, respectively

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL: 09), Base Year, Student Survey, 2009

Table 10 Reliability for students' feelings towards their math and science classroom teachers

Items	Strongly disagree			Strongly agree
1. Math/science teacher values and listens to students' ideas	1	2	3	4
2. Math/science teacher treats students with respect	1	2	3	4
3. Math/science teacher treats every student fairly	1	2	3	4
4. Math/science teacher thinks every student can be successful	1	2	3	4
5. Math/science teacher thinks mistakes are okay as long as all students learn	1	2	3	4
6. Math/science teacher treats some kids better than other kids	1	2	3	4
7. Math/science teacher makes math/science interesting	1	2	3	4
8. Math/science teacher treats males and females differently	1	2	3	4
9. Math/science teacher makes math easy to understand	1	2	3	4
Reliability	Cronbach's Alpha (α)			
	Feeling about math teacher		Feeling about science teacher	
	0.89		0.89	

Originally, Item 6 and Item 8 are negatively worded, all other items are coded as 1 = strongly agree, 4 = strongly disagree. In this study, we recoded these as 1 = strongly disagree, 4 = strongly agree. Due to the missing data, 18,880 and 17,420 observations were tested for students' feeling about math and science teachers, respectively

Source: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSL: 09), Base Year, Student Survey, 2009

Acknowledgements

Earlier versions of this manuscript benefitted from formative feedback at the American Sociological Association Annual Meeting, and has benefitted from guidance from other members of our internal research colleagues, in particular the other members of Dr. Zhao's dissertation committee: Drs. Frances Berry, Christine Mokher, Toby Park-Gaghan, and Yanyun Yang.

Author contributions

TZ led the manuscript in its entirety, an extension from his doctoral dissertation, and developed the drafts and analytic design. P-F procured the data, contributed to the analytic approach, reviewed the analyses, led the revisions, and contributed sections and edits to the manuscript. All authors read and approved the final manuscript.

Funding

This material is based upon work supported by the National Science Foundation under Grant No. 1920670. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Availability of data and materials

Public use access to these data is available through the National Center for Education Statistics website, as is the application for restricted-use data, such as we used for these analyses: https://nces.ed.gov/surveys/hsls09/hsls09_data.asp. Our restricted-use license number is 12100041. Statistical code generated to analyze supporting the findings of this study is available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate

The project's data use and design were approved by the Florida State University Institutional Review Board #2019.28616.

Consent for publication

Not applicable—deidentified large-scale database analysis.

Competing interests

We have no conflicts to disclose.

Author details

¹Chinese Academy of Science and Education Evaluation, Hangzhou Dianzi University, Hangzhou, China. ²Department of Educational Leadership and Policy Studies and Center for Postsecondary Success, Florida State University, 1114 West Call Street, Tallahassee, FL 32306, USA.

Received: 22 September 2021 Accepted: 3 June 2022

Published online: 22 June 2022

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135. <https://doi.org/10.1086/508733>
- Allen, D., Dancy, M., Stearns, E., Mickelson, R., & Bottia, M. (2022). Racism, sexism and disconnection: Contrasting experiences of Black women in STEM before and after transfer from community college. *International Journal of STEM Education*. <https://doi.org/10.1186/s40594-022-00334-2>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117–148. https://doi.org/10.1207/s15326985sep2802_3
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72(1), 187–206. <https://doi.org/10.1111/1467-8624.00273>
- Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59(5), 960–970. <https://doi.org/10.1037/0022-3514.59.5.960>
- Bobbitt-Zeher, D. (2007). The gender income gap and the role of education. *Sociology of Education*, 80(1), 1–22. <https://doi.org/10.1177/003804070708000101>
- Bottia, M. C., Stearns, E., Mickelson, R. A., & Moller, S. (2018). Boosting the numbers of STEM majors? The role of high schools with a STEM program. *Science Education*, 102(1), 85–107. <https://doi.org/10.1002/sce.21318>
- Bottia, M. C., Stearns, E., Mickelson, R., Moller, S., & Parker, A. D. (2015). The relationships among high school STEM learning experiences and students' intent to declare and declaration of a STEM Major in college. *Teachers College Record*, 117(4), 1–46. <https://doi.org/10.1177/016146811511700308>
- Carter, D. F., Razo Dueñas, J. E., & Mendoza, R. (2019). Critical examination of the role of STEM in propagating and maintaining race and gender disparities. In M. B. Paulsen & L. W. Perna (Eds.), *Higher education: handbook of theory and research* (Vol. 34, pp. 39–97). Springer International Publishing. https://doi.org/10.1007/978-3-030-03457-3_2
- Corbett, C., & Hill, C. (2015). *Solving the Equation: The Variables for Women's Success in Engineering and Computing*. American Association of University Women. <http://www.aauw.org/files/2015/03/Solving-the-Equation-report-nsa.pdf>.
- Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *The American Journal of Sociology*, 106(6), 1691–1730. <https://doi.org/10.1086/321299>
- Crisp, G., Nora, A., & Taggart, A. (2009). Student characteristics, pre-college, college, and environmental factors as predictors of majoring in and earning a STEM degree: An analysis of students attending a Hispanic-serving institution. *American Educational Research Journal*, 46(4), 924–942. <https://doi.org/10.3102/0002831209349460>
- Csikszentmihályi, M., Rathunde, K., & Whalen, S. (1993). *Talented teenagers: The roots of success and failure*. Cambridge University Press.
- Dalton, B., Ingels, S. J., Downing, J., & Bozick, R. (2007). *Advanced Mathematics and Science Course-taking in the Spring High School Senior Classes of 1982, 1992, and 2004* (NCES 2007-312). <http://nces.ed.gov/pubs2007/2007312.pdf>.
- Dika, S. L., & D'Amico, M. M. (2016). Early experiences and integration in the persistence of first-generation college students in STEM and non-STEM majors. *Journal of Research in Science Teaching*, 53(3), 368–383. <https://doi.org/10.1002/tea.21301>
- Doren, C., & Lin, K. Y. (2019). Diverging Trajectories or Parallel Pathways? An intersectional and life course approach to the gender earnings gap by race and education. *Socius*. <https://doi.org/10.1177/2378023119873816>
- Duprey, M.A., Pratt, D.J., Jewell, D.M., Cominole, M.B., Fritch, L.B., Ritchie, E.A., Rogers, J.E., Wescott, J.D., Wilson, D.H. (2018). *High School Longitudinal Study of 2009 (HSL:09) Base-Year to Second Follow-Up Data File Documentation* (NCES 2018-140). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from <https://nces.ed.gov/pubsearch>.
- Eccles, J. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). W. H. Freeman.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and socio-cultural perspective on motivation. *Contemporary Educational Psychology*, 61, 101859.
- Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., Hurtado, S., John, G. H., Matsui, J., McGee, R., Okpodu, C. M., Robinson, T. J., Summers, M. F., Werner-Washburne, M., & Zavala, M. (2016). Improving underrepresented minority student persistence in STEM. *CBE Life Sciences Education*, 15(3), e5. <https://doi.org/10.1187/cbe.16-01-0038>
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole Publishing.
- Fayer, S., Lacey, A., & Watson, A. (2017). *STEM occupations: Past, present, and future. Spotlight on statistics*. U.S. Department of Labor, Bureau of Labor Statistics. Retrieved from: https://digitalcommons.ilr.cornell.edu/key_workplace/1923/.

- Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, 20(2), 507–537. <https://doi.org/10.1111/j.1532-7795.2010.00645.x>
- Garrison, H. (2013). Underrepresentation by race-ethnicity across stages of U.S. science and engineering education. *CBE—Life Sciences Education*, 12(3), 357–363. <https://doi.org/10.1187/cbe.12-12-0207>
- Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review*, 29(6), 911–922. <https://doi.org/10.1016/j.econedurev.2010.06.010>
- Harro, B. (2000). The cycle of socialization. In M. Adams, W. J. Blumenfeld, R. Castañeda, H. W. Hackman, M. L. Peters, & X. Zuñiga (Eds.), *Readings for diversity and social justice: An anthology on racism, antisemitism, sexism, heterosexism, ableism, and classism* (pp. 15–21). Routledge.
- Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, 70(2), 151–179. <https://doi.org/10.3102/00346543070002151>
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4
- Huang, G., Taddese, N., & Walter, E. (2000). Entry and persistence of women and minorities in college science and engineering education (NCES 2000-601). National Center for Educational Statistics. Retrieved from: https://nces.ed.gov/programs/quarterly/vol_2/2_3/post_women.asp.
- Ingels, S. J., Pratt, D. J., Herget, D. R., Dever, J. A., Fritch, L. B., Ottem, R., Rogers, J. E., Kitmitto, S., & Leinwand, S. (2014). *High School Longitudinal Study of 2009 (HLS: 09) Base Year to First Follow-Up Data File Documentation*. National Center for Education Statistics. Retrieved from: <https://eric.ed.gov/?id=ED565693>.
- Kaufman, R. L. (2018). *Interaction effects in linear and generalized linear models*. Sage.
- Kemp, A. H., Quintana, D. S., Gray, M. A., Felmingham, K. L., Brown, K., & Gatt, J. M. (2010). Impact of depression and antidepressant treatment on heart rate variability: A review and meta-analysis. *Biological Psychiatry*, 67(11), 1067–1074.
- Ketenci, T., Leroux, A., & Renken, M. (2020). Beyond student factors: A study of the impact on STEM career attainment. *Journal for STEM Education Research*, 3(3), 368–386. <https://doi.org/10.1007/s41979-020-00037-9>
- Kline, P. (2013). *Handbook of psychological testing*. Routledge.
- Lee, V. E., Chow-Hoy, T., Burkam, D. T., Gevertz, D., & Smerdon, B. A. (1998). Sector differences in high school course taking: A private school or Catholic school effect? *Sociology of Education*, 71(4), 314–335. <https://doi.org/10.3102/00028312036004907>
- Legewie, J., & DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociology of Education*, 87(4), 259–280. <https://doi.org/10.1177/0038040714547770>
- Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*, 95(5), 877–907. <https://doi.org/10.1002/sce.20441>
- Martinot, D., & Désert, M. (2007). Awareness of a gender stereotype, personal beliefs, and self-perceptions regarding math ability: When boys do not surpass girls. *Social Psychology of Education*, 10(4), 455–471. <https://doi.org/10.1007/s11218-007-9028-9>
- Milesi, C., Perez-Felkner, L., Brown, K. L., & Schneider, B. (2017). Engagement, persistence, and gender in computer science: Results of a smartphone ESM study. *Frontiers in Psychology*, 8(602), 1–9. <https://doi.org/10.3389/fpsyg.2017.00602>
- Niu, L. (2017). Family socioeconomic status and choice of STEM major in college: An analysis of a national sample. *College Student Journal*, 51(2), 298–312.
- Nix, S., & Perez-Felkner, L. (2019). Difficulty orientations, gender, and race/ethnicity: An intersectional analysis of pathways to STEM degrees. *Social Sciences*, 8(2), 1–29. <https://doi.org/10.3390/socsci8020043>
- Oguzoglu, U., & Ozbeklik, S. (2016). *Like father, like daughter (unless there is a son): Sibling sex composition and women's STEM major choice in college*. IZA Discussion Paper No. 10052. Retrieved from <https://ssrn.com/abstract=2810463>.
- Pais, J. (2011). Socioeconomic background and racial earnings inequality: A propensity score analysis. *Social Science Research*, 40(1), 37–49. <https://doi.org/10.1016/j.sresresearch.2010.06.016>
- Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology*, 106(1), 315. <https://doi.org/10.1037/a0034027>
- Perez-Felkner, L., McDonald, S.-K., & Schneider, B. L. (2014). What happens to high-achieving females after high school? Gender and persistence on the postsecondary STEM pipeline. In I. Schoon & J. S. Eccles (Eds.), *Gender differences in aspirations and attainment: A life course perspective* (pp. 285–320). Cambridge University Press. <https://doi.org/10.1017/CBO9781139128933.018>
- Perez-Felkner, L., Nix, S., & Thomas, K. (2017). Gendered pathways: How mathematics ability beliefs shape secondary and postsecondary course and degree field choices. *Frontiers in Psychology*, 8(386), 1–11. <https://doi.org/10.3389/fpsyg.2017.00386>
- Perez-Felkner, L., Thomas, K., Nix, S., Hopkins, J., & D'Sa, M. (2019). Are 2-year colleges the key? Institutional variation and the gender gap in undergraduate STEM degrees. *The Journal of Higher Education*, 90(2), 181–209. <https://doi.org/10.1080/00221546.2018.1486641>
- Phelps, L. A., Camburn, E. M., & Min, S. (2018). Choosing STEM college majors: Exploring the role of pre-college engineering courses. *Journal of Pre-College Engineering Education Research (j-PEER)*, 8(1), 1. <https://doi.org/10.7771/2157-9288.1146>
- Quadlin, N. (2017). Funding Sources, family income, and fields of study in college. *Social Forces*, 96(1), 91–120. <https://doi.org/10.1093/sf/sox042>
- Renninger, K. A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist*, 46(3), 168–184. <https://doi.org/10.1080/00461520.2011.587723>
- Renninger, K. A., Hidi, S., & Krapp, A. (Eds.). (1992). *The role of interest in learning and development*. Lawrence Erlbaum Associates.
- Riegle-Crumb, C., & King, B. (2010). Questioning a white male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Educational Researcher*, 39(9), 656–664. <https://doi.org/10.3102/0013189x10391657>
- Riegle-Crumb, C., King, B., & Irizarry, Y. (2019). Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields. *Educational Researcher*, 48(3), 133–144. <https://doi.org/10.3102/0013189x19831006>
- Riegle-Crumb, C., Morton, K., Nguyen, U., & Dasgupta, N. (2019). Inquiry-based instruction in science and mathematics in middle school classrooms: Examining its association with students' attitudes by gender and race/ethnicity. *AERA Open*, 5(3), 1–17. <https://doi.org/10.1177/2332858419867653>
- Riegle-Crumb, C., & Peng, M. (2021). Examining high school students' gendered beliefs about math: Predictors and implications for choice of STEM college majors. *Sociology of Education*, 94(3), 227–248. <https://doi.org/10.1177/00380407211014777>
- Saw, G., Chang, C.-N., & Chan, H.-Y. (2018). Cross-sectional and longitudinal disparities in STEM career aspirations at the intersection of gender, race/ethnicity, and socioeconomic status. *Educational Researcher*, 47(8), 525–531. <https://doi.org/10.3102/0013189x18787818>
- Schiefele, U. (2009). Situational and individual interest. In K. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 197–222). Routledge.
- Schneider, B., Broda, M., Judy, J., & Burkander, K. (2013). Pathways to college and STEM careers: Enhancing the high school experience. *New Directions for Youth Development*, 2013(140), 9–29. <https://doi.org/10.1002/yd.20076>
- Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Broda, M., Spicer, J., Bruner, J., Moeller, J., Linnansaari, J., Juuti, K., & Viljaranta, J. (2016). Investigating optimal learning moments in U.S. and Finnish science classes. *Journal of Research in Science Teaching*, 53(3), 400–421. <https://doi.org/10.1002/tea.21306>
- Scott, K. A., Sheridan, K. M., & Clark, K. (2015). Culturally responsive computing: A theory revisited. *Learning, Media and Technology*, 40(4), 412–436. <https://doi.org/10.1080/17439884.2014.924966>
- Shapiro, C. A., & Sax, L. J. (2011). Major selection and persistence for women in STEM. *New Directions for Institutional Research*, 2011(152), 5–18. <https://doi.org/10.1002/ir.404>
- St. John, E. P., & Asker, E. H. (2003). *Refinancing the college dream: Access, equal opportunity, and justice for taxpayers*. Johns Hopkins University Press.
- Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*, 39(2), 81–90. <https://doi.org/10.1080/01973533.2016.1277529>

- 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. <https://doi.org/10.1080/10824660701601266>
- Wagner, M., Marder, C., & Blackorby, J. (2002). *The Children We Serve: The demographic characteristics of elementary and middle school students with disabilities and their households*. SRI International.
- Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, 24(5), 770–775. <https://doi.org/10.1177/0956797612458937>
- Wang, X. (2013a). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, 50(5), 1081–1121. <https://doi.org/10.3102/0002831213488622>
- Wang, X. (2013b). Modeling entrance into STEM fields of study among students beginning at community colleges and four-year institutions. *Research in Higher Education*, 54(6), 664–692. <https://doi.org/10.1007/s11162-013-9291-x>
- Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: A comparison of samples from Australia, Canada, and the United States. *Developmental Psychology*, 48(6), 1594–1611. <https://doi.org/10.1037/a0027838>
- Wiswall, M., & Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2), 791–824. <https://doi.org/10.1093/restud/rdu044>
- Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. *Annual Review of Sociology*, 41(1), 331–357. <https://doi.org/10.1146/annurev-soc-071312-145659>
- Xu, Y. (2015). Focusing on women in STEM: A longitudinal examination of gender-based earning gap of college graduates. *The Journal of Higher Education*, 86(4), 489–523. <https://doi.org/10.1080/00221546.2015.11777373>
- Xu, Y. (2016). Aspirations and application for graduate education: Gender differences in low-participation STEM disciplines. *Research in Higher Education*, 57(8), 913–942. <https://doi.org/10.1007/s11162-016-9411-5>
- Zhang, L. (2008). Gender and racial gaps in earnings among recent college graduates. *The Review of Higher Education*, 32(1), 51–72. <https://doi.org/10.1353/rhe.0.0035>
- Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663–676. <https://doi.org/10.3102/00028312029003663>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
