Econ 496S

Bridging the Persistence Gap: An Investigation of the Underrepresentation of Female and Minority Students in STEM Fields

Aaditya Jain and Bailey Kaston

Peter Arcidiacono, Faculty Advisor

Michelle Connolly, Faculty Advisor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University
Durham, North Carolina
April 20, 2020

Abstract

Prior literature on mismatch theory has concentrated primarily on minority students, whose lower average levels of pre-enrollment preparedness tend to discourage them from persisting in STEM fields as often as their non-minority counterparts at selective universities. Our study shifts the focus to the persistence gap between men and women, invoking social cognitive career theory to investigate how factors beyond preparedness – such as self-confidence – cause women to switch out of selective STEM programs at higher rates than men. Using the High School Longitudinal Study of 2009, we investigate the drivers of STEM persistence for all students and arrive at two main conclusions. First, higher levels of STEM preparedness are more beneficial to STEM persistence at selective universities, confirming mismatch theory in the sample. We then simulate the counterfactual scenario and find that 33% of students at selective schools would have been more likely to persist in STEM had they attended less selective schools, a figure that reaches 50% for underconfident female students. This observation ties to our second conclusion – that underconfidence in math relative to one's true performance decreases the likelihood of STEM persistence for all students at selective universities, and that female students at selective schools are more likely to be underconfident than their male counterparts. Our findings suggest that the appropriate policy solution to reduce STEM attrition rates among women should then become a two-pronged approach: (1) more selective universities should better support the STEM self-confidence levels of female students, and (2) home environments should ideally cultivate that self-confidence long before women even reach college. In our final set of analyses, we thus explore the factors that drive math overconfidence in the first place, and conclude that both student and parental biases against female STEM ability are detrimental to the STEM self-confidence of female students.

Keywords: Education; STEM; Mismatch Theory; Social Cognitive Career Theory

JEL Classification: I24, I26, J24

Introduction

Careers in science, technology, engineering, and mathematics (STEM) are essential to the sustained innovative and intellectual competitiveness of the United States (Noonan, 2017), and individuals who work in STEM fields have long been compensated with an earnings premium (Arcidiacono, 2004; Hastings, Neilson, & Zimmerman, 2013; Kirkeboen, Leuven, & Mogstad, 2016). A 2019 Wall Street Journal compilation of data from the United States Education Department showed that STEM majors earn higher median salaries than students from any other academic concentration in the year immediately following college graduation (Mitchell, Fuller, & Hackman, 2019). Higher compensations in STEM are driven largely by a growing demand for STEM talent in the workforce – while non-STEM employment opportunities have grown a meager 4.0 percent since 2005, employment in STEM fields has grown by 24.4 percent over the same period (Noonan, 2017). The STEM sector is also expected to continue to expand in the future, with over half a million projected new STEM jobs opening in the United States by the end of 2024 (Fayer, Lacey, & Watson, 2017).

However, despite the societal importance, financial rewards, and unrivaled growth of STEM employment, overall student interest in STEM career paths has remained largely the same over time. For instance, an annual survey administered by the Higher Education Research Institute (HERI) at the University of California–Los Angeles (UCLA) revealed that the percentage of incoming college students with an initial preference for a STEM major has hovered rather steadily around 31 percent since 1971 (Carnevale, Smith, & Melton, 2011). Thus, the United States economy is predicted to endure a shortage of STEM professionals in the future, prompting researchers and policymakers to investigate and devise policies to increase educational investments and degree completion in STEM. Prior studies have explored avenues as distinct as making financial aid more accessible (Castleman, Long, & Mabel, 2017) and reducing class sizes in primary schools (Dynarski, Hyman, & Schanzenbach, 2013). The impending shortage of STEM professionals has also yielded an even more direct potential solution: attempting to decrease the attrition and dropout of students from STEM fields at the college level. For example, a 2012 report by the White House argued that increasing the national persistence rate in STEM by only 10% would reduce the shortage of STEM professionals in 2022 by 75%. In this paper, we expound upon this avenue by analyzing the factors that drive differences in persistence rates along racial/gender lines and explain the underrepresentation of minority and female students in college STEM programs.

As STEM interest has stagnated over time, racial gaps between minority and non-minority populations have persisted across all levels of society – academic achievement, professional life, and economic outcomes. At the academic level, several studies have demonstrated that, despite improvements over the past few decades, there remain significant achievement gaps that originate as early as elementary school (Fryer and Levitt, 2006) and tend to persist throughout high school (Hanushek and Rivkin, 2008), college (Arcidiacono et al., 2012), and even graduate school (Bleske-Rechek and Browne, 2014) for minority groups. Most notably for our study, there is a substantial racial divide in STEM fields at the college level, as minority students have much lower persistence rates in STEM and are far less likely than their non-minority peers to graduate with a STEM degree despite indicating similar levels of interest in STEM prior to enrollment in college (Anderson & Kim, 2006; Arcidiacono et al., 2012).

Parallel to the racial persistence gap that exists in STEM studies, there also exists a stark disparity in STEM persistence rates between men and women. Prior studies have demonstrated that women comprise a disproportionately low proportion of the STEM degree-holding population, despite making up nearly half of the United States' college-educated workforce (Beede et al., 2011; Card & Payne, 2017). The underrepresentation of women in math-intensive STEM fields begins to develop as early as high school advanced math classes, and only continues to persist through college and graduate school (Wang & Degol, 2017). Although the STEM gender gap has narrowed somewhat in recent years, this pattern is driven more by an increase in the number of women entering higher education institutions rather than a change in females' preferences and persistence in STEM – women still demonstrate preferences for non-STEM fields at significantly higher rates than their male counterparts and have a lower rate of persistence in STEM fields than their male counterparts, positing the STEM gender gap as an important social policy issue (Mann & DiPrete, 2013).

The underrepresentation of women and minority individuals in STEM fields imposes tangible and significant economic drawbacks. Most notably, divergence in STEM representation is a key driver in the wage gap between those underrepresented groups and the traditional male, Caucasian archetype (Card & Payne, 2017). Women in STEM occupations earn 33% more than women in non-STEM occupations, while the STEM wage premium for men is 25% (Beede et al., 2011). Even after controlling for similar educational achievements, wage outcomes tend to favor those in STEM occupations. Thus, underrepresentation in STEM for women and minority individuals perpetuates inequality across all levels of society. Thus, a more comprehensive understanding of the gap between student groups who more consistently persist in STEM and those who do not could allow us to more thoughtfully enact policies and establish programs that boost degree completion for underrepresented minorities and females in STEM.

Specifically, using a nationally representative dataset that tracks 23,000+ students from 9th grade through their junior year of college, we hope to identify why racial and gender disparities exist in persistence of STEM majors at the college level, with minority and female students majoring in STEM fields and graduating with STEM degrees at significantly lower rates than their Caucasian and Asian male counterparts.

The Persistence Gap

As incoming college freshmen, African American and Hispanic students express initial preferences for STEM majors at nearly identical rates to white and Asian students. Nevertheless, the probability that those black or Hispanic students graduate with their targeted STEM degrees is far lower than that of their white and Asian peers (Anderson & Kim, 2006; Arcidiacono, Aucejo, & Hotz, 2013). In response to these concerning trends at universities, other researchers have attempted to find out what drives the persistence gap between demographic groups; in other words, why do minority students switch out of STEM majors (Arcidiacono, Aucejo, & Spenner, 2012) – or drop out of school altogether (Luppino & Sander, 2013) – at higher rates than non-minority students? By studying two Duke University cohorts, Arcidiacono et al. (2012) attribute the different STEM switching rates to the differences in academic preparation between minority and non-minority students before college. They find that, after controlling for SAT score as a proxy for pre-enrollment college preparedness, the difference in STEM switching rates between minority and non-minority students becomes negligible. So, race itself is *not* the driving factor behind the persistence gap between minority and non-minority students; rather, the lower levels of pre-enrollment academic preparation among minority students more commonly result in major switching decisions as a reaction to the typically tougher grading standards and time requirements of a STEM course load. In a separate study, Arcidiacono, Aucejo, and Hotz (2013) utilize rich student-level data from the University of California system to once more demonstrate that pre-college academic preparedness drives major choices; less prepared students tend to both switch out of science majors with higher frequencies and take longer to graduate from their respective universities. In a similar vein, Stinebrickner and Stinebrickner (2012, 2013) conduct two separate studies utilizing an institution-specific Berea Panel Study dataset. Their results across these two studies suggest that minority students who enter college open to science majors shift away from STEM fields after realizing that their grade performance will be significantly lower than expected. The initial relative optimism of students studying in a STEM field is tempered by corrections to performance expectations once the university experience begins.

As for the gender persistence gap, women both show lower interest in STEM and persist in their STEM majors at lower rates than men (Kanny et al., 2014). Additionally, women who do pursue STEM fields tend to disproportionately gear their studies towards the biological sciences rather than the physical sciences and engineering; NCES data show that in the last 25 years, women have only marginally increased their share of engineering degrees, while drastically decreasing their share of computer science degrees (Kanny et al., 2014).

Prior research has culminated in three dominant lines of thinking regarding the gender gap: gender essentialism, gender environmentalism, and gender constructivism (Liben & Coyle, 2014). Gender essentialism postulates that gender is essential to human nature and pinpoints biology, including evolutionary theory, as the key driver of behavioral outcomes. Fundamentally, gender essentialism argues that it is futile to intervene in gender differentiation. However, gender essentialism is not necessarily defeatist, as some gender essentialist theories argue that because the two genders are profoundly different, it makes sense to close the gender gap by approaching female interest and persistence in STEM differently than we approach male participation in STEM (Liben & Coyle, 2014). Researchers such as Liben & Coyle (2014) often propose singlesex public education as a pathway to higher female STEM participation. Gender environmentalism, by contrast, emphasizes the impact of students’ surroundings on their behavioral outcomes. How one’s parents, teachers, and peers approach STEM and how they perceive the student’s STEM abilities are particularly important here. Lastly, rather than assigning explanatory power explicitly to nature or nurture, gender constructivism emphasizes the role of individuals in shaping their own behavioral outcomes. One branch of gender constructivism, social cognitive career theory (SCCT), argues that three key factors – personal perception, outcome expectations, and goals – drive a life-long process of developing career aspirations, with college major choices encapsulating a benchmark achievement in that path (Lent, Brown, and Hackett, 1994). SCCT suggests that students’ experiences dictate their self-perception of their capabilities and that this notion of self-efficacy dictates entrance and persistence in STEM. For underrepresented groups in STEM, institutional barriers such as sexism and racism can be monumental obstacles to overcome due to their effects on personal perception and outcome expectations (Fouad & Santana, 2017). Similarly, prior studies have proposed that the supposed “masculinity” of STEM subjects such as math and physics may also act as an additional barrier for females in STEM (Makarova et al., 2019).

Selectivity of the University

Given the supporting evidence that pre-enrollment academic preparation is a key determinant in driving students’ decisions to switch out of STEM fields for both female and minority students, the literature has shifted its focus to the differences between STEM student behavior at more selective and less selective universities. Racial and gender preferences in admissions policies at selective universities, engendered both by a drive towards campus diversity and affirmative action policies, are thought to have the effect of placing a large proportion of high-potential, under-prepared students in rigorous STEM environments. *Mismatch theory* suggests that because minority students often come from less privileged backgrounds with fewer resources at their disposal and lower levels of pre-enrollment academic preparation, they ultimately find it more difficult to succeed in competitive STEM environments. In their two studies of the UC system, Arcidiacono et al. (2012, 2013) find that this type of mismatch occurs in that minority students with similar observable traits graduate at higher rates and in fewer semesters in STEM fields when attending less selective UC colleges than their similar counterparts at more selective UC colleges. Additionally, Arcidiacono et al. (2013) show that after California implemented Proposition 209, which effectively banned affirmative action policies at the state's public universities, the distribution of minority enrollment shifted towards less selective colleges. Consequently, overall minority persistence in STEM majors increased due to the higher enrollment at less selective universities.

Collectively, these studies suggest that more selective universities are better at educating and graduating the more prepared students in STEM fields, while less selective schools are better at educating and graduating less prepared STEM students. Smyth and McArdle (2004), who look at 23 unique selective universities, similarly find that attending an elite university does not increase the chances of completing a science, engineering, or math major after controlling for pre-enrollment preparation. Thus, there is prior supporting evidence of the existence of mismatch at U.S. universities. Identifying whether mismatch exists between selective colleges and high-potential minority and female students is extremely important to assess whether admissions policies are improving minority and gender representation in STEM professions, or whether these policies are instead hindering the ability of underrepresented students to pursue STEM education by placing students in STEM environments where they are less likely to succeed.

Although there is already an extensive literature base on the difference between STEM switching rates among minority students across universities with varying levels of selectivity, the 944 high schools and 10 states represented in our dataset provide us the opportunity to conduct the first nationally representative study of this kind. Additionally, the dataset we use is specifically geared towards answering questions about STEM education in the United States. The dataset is rich, not only in the sense of traditional observable characteristics for students and their backgrounds, but also in the sense that it captures intangible and traditionally unobservable traits such as student and parent expectations and preferences and ethereal qualities such as motivation and peer effects. By capturing these various factors, we can more effectively investigate the elusive counterfactual scenario and observe how STEM persistence levels differ among students who are similar in all facets except the selectivity of the university they attend.

High School Longitudinal Study (HSLS) of 2009

To advance prior research and explore potential causes for the persistence gap for female and minority students, we will employ the High School Longitudinal Study (HSLS) of 2009, one of several cohort studies conducted by the National Center for Education Statistics (NCES). The data collected are nationally representative, following 23,000+ randomly selected 9th graders starting in 2009 from 944 different high schools, with an average of 25 students per school. There are follow-ups in 2012 when students are high school seniors, in 2013 to collect secondary school transcripts before students enter college, and again in 2016 when students are in their junior year of college and have declared their intended field of study. There is an additional planned follow-up in 2025 to survey students' post-education economic outcomes. Table 1 details the sample distribution of all students who fall into the four most concentrated racial categories in the dataset.

Table 1: Sample population by gender and race/ethnicity

Gender and Race		Frequency	Percentage
Female	White	6,357	30.36
	Black	1,276	6.09
	Hispanic	1,796	8.59
	Asian	970	4.63
Male	White	6,594	31.51
	Black	1,172	5.59
	Hispanic	1,816	8.68
	Asian	952	4.55
Total		20,933	100.00

Source: High School Longitudinal Study of 2009

Table 2 presents key student characteristics by race and gender for the whole dataset where responses were available. The first row, gender, shows that across our four main racial categories, the sample distribution is relatively even, with near-50/50 splits. The second row illustrates the well-documented fact in the United States that minorities disproportionately fall into lower-income groups, while white and Asian individuals dominate the higher income brackets. As shown in the table, nearly half of the white and Asian students in the sample belong to the top two income quintiles, while more than half of the black students and more than 60% of Hispanic students fall within the bottom three income quintiles. This reaffirms our understanding that minority students tend to have disproportionately fewer resources, which ties to the third row of Table 2 regarding standardized math test scores. Here, we see that black and Hispanic students (who, on average, come from lower-income families) tend to disproportionately fall in the lower quintiles of test scores, while white and Asian students (who, on average, come from higher-income families) dominate the top-scoring quintiles.

Table 2: Summary statistics for selected variables by race

Summary Characteristics		White	Black	Hispanic	Asian
Gender	Female	0.491	0.479	0.503	0.495
	Male	0.509	0.521	0.497	0.505
Family Income	First Quintile	0.109	0.220	0.323	0.131
	Second Quintile	0.164	0.202	0.222	0.128
	Third Quintile	0.186	0.211	0.179	0.145
	Fourth Quintile	0.239	0.198	0.149	0.208
	Fifth Quintile	0.302	0.169	0.126	0.388
Standardized Mathematics Test Score	First Quintile	0.142	0.281	0.210	0.056
	Second Quintile	0.165	0.243	0.231	0.093
	Third Quintile	0.190	0.212	0.223	0.131
	Fourth Quintile	0.235	0.169	0.193	0.212
	Fifth Quintile	0.268	0.095	0.144	0.508

Source: High School Longitudinal Study of 2009

Table 3 recreates Table 2 but focuses only on the subset of the students that enter college with an initial STEM major. Here we see that the gender distribution skews severely towards males across all races. Even for initial STEM college majors, there is a replicated trend from Table 2 that minorities disproportionately fall into the first three income brackets while non-minorities dominate the top two income brackets. Again, we observe that this lack of resources for minority groups also tends to translate into lower standardized math scores. Interestingly, the proportion of Hispanic students across the top two quintiles is about the same as white students, albeit with a greater proportion of white students in the top quintile.

Table 3: Summary statistics for selected variables by race for initial STEM major

Summary Characteristics		White	Black	Hispanic	Asian
Gender	Female	0.352	0.423	0.363	0.420
	Male	0.648	0.577	0.637	0.580
Family Income	First Quintile	0.029	0.081	0.223	0.074
	Second Quintile	0.073	0.158	0.153	0.065
	Third Quintile	0.132	0.189	0.164	0.115
	Fourth Quintile	0.220	0.248	0.162	0.166
	Fifth Quintile	0.468	0.248	0.244	0.481
Standardized Mathematics Test Score	First Quintile	0.029	0.061	0.054	0.009
	Second Quintile	0.060	0.149	0.100	0.032
	Third Quintile	0.104	0.223	0.146	0.057
	Fourth Quintile	0.225	0.302	0.245	0.171
	Fifth Quintile	0.581	0.265	0.460	0.731

Source: High School Longitudinal Study of 2009

The HSLS not only conducts student questionnaires, but also surveys students' parents, math and science teachers, school administrators, and guidance counselors. Additionally, the study includes a standardized math assessment both in the fall of 9th grade and the spring of 11th grade. This comprehensive approach provides a more holistic understanding of what drives students' decision-making with regards to STEM studies. The HSLS as a whole was engineered to understand how policy innovation can impact the STEM pipeline for college and beyond and give policymakers a better understanding of how to target underrepresented STEM groups most effectively. Specifically, the HSLS uses the student and parent surveys to gauge otherwise unobservable characteristics such as student expectations, parent expectations, and student work ethic. As a result, this dataset will allow us to study the persistence gap as more than just a function of observable student preparedness proxied by test scores. The survey of students at each checkpoint includes questions like:

1. "How confident are you that you can do an excellent job on tests in this [math] course?"

2. “How many of your close friends get good grades? Plan to attend a 4-year college?”
3. “On a typical weekday during the school year, how many hours do you spend working on math homework and studying for math tests?”

In the base year, the survey similarly asks parents:

1. “If there were no barriers, how far in school would you want [your 9th grader] to go?”
2. “As things stand now, how far in school do you think [he/she] will actually get?
3. “In general, how would you compare males and females in math?”

These expectation and preference questions are important to get better insight into the true driving factors of STEM behavior across demographics. For instance, perhaps female students are more likely to switch out of STEM majors than their male counterparts because of societal expectations and peer beliefs about the ‘typical’ occupation of a woman. Or, perhaps selectivity of the university and overconfidence of students relative to their actual performance mediate the severity of the persistence gap between underrepresented and overrepresented students.

Empirical Framework, Results, & Discussion

1. Understanding Basic Differences in Persistence Rates

We begin by focusing our attention on racial and gender disparities in persistence, as well as quantitative observable student characteristics such as pre-enrollment academic preparedness. Our baseline regression is our simplest, using the dataset’s standardized math test score as a proxy for preparedness and a composite socioeconomic metric that captures parents’ education, occupation, and income. We then delve into the idea of over/underconfidence relative to objective math ability and the role it plays in boosting and/or diminishing persistence rates. Note that our sample size for the following regressions will narrow from the original 22,954 students presented in tables 1 and 2 to only those students who attend college and initially plan on majoring in STEM fields. Additionally, our sample size will fluctuate slightly in the following regressions due to missing or non-response answers for some of the HSLS survey questions and math assessments.

1.1 Baseline Regression

The main variable of interest in our baseline regression is students' standardized math scores, derived from an assessment issued by the HSLS. As noted earlier, several studies – such as Arcidiacono's on racial differences in GPA (2012) – have concluded that overall academic preparedness is a major driver of persistence in STEM. The results from Equation 1.1, shown in column 1 of Figure 1 below, interestingly defy our initial hypothesis that *Selectivity* should be a significant negative predictor of persistence rates. Furthermore, *Minority* – a dummy that indicates whether a student is black or Hispanic – is also not significant. Instead, we initially see that both *MathScore* and *Female* are significant. *Female* is significant with a negative coefficient, a result that has been observed in the prior literature as women tend to persist at significantly lower rates.

$$1.1\ Pr(Persist\ in\ STEM) = \Phi(\beta_1 University\ Selectivity + \beta_2 Female + \beta_3 Minority + \beta_4 Socioeconomic\ Composite + \beta_5 Math\ Score)$$

1.2 Role of Confidence in Persistence

Given that our first-pass regression illustrates the importance of our academic preparedness proxy (math score), while seemingly challenging the idea of mismatch theory, we work to identify the attitudinal and environmental factors that may contribute to persistence. The first attitudinal factor that we tackle is that of math confidence. The dataset provides us with a proxy for self-confidence in math that is based on a composite of students' self-reported performance on school math tests, comfort levels with their math textbook, and confidence in their abilities to do an excellent job on math homework assignments. Table 4 presents a breakout, by gender and race, of math score, math confidence, and STEM persistence rates for only those students who initially declared a major in STEM.

Table 4: Breakout of Gender and Race Against Math Achievement and STEM Studies

Gender and Race		Math Score	Math Confidence	Persisted in STEM	Frequency
Female	White	59.85	0.337	0.693	596
	Black	54.37	0.376	0.688	93
	Hispanic	57.35	0.326	0.728	122
	Asian	63.68	0.422	0.781	225
Male	White	59.91	0.535	0.771	1,103
	Black	52.97	0.662	0.752	122
	Hispanic	56.67	0.572	0.777	213
	Asian	64.57	0.575	0.831	306

Source: High School Longitudinal Study of 2009

Table 4 highlights two very interesting trends. First, we see that across genders, there is a stark contrast in math confidence despite similar – if not higher – math scores for women. For instance, there is a 25-30 percentage point difference between minority male and female confidence levels even though minority women tend to perform at least 1-2 percentage points higher on the standardized math assessment. Secondly, women across all racial groups persist in

STEM at much lower rates despite their greater performance in math. The table thus demonstrates that rather than math score, confidence and self-esteem may be the best indicators of STEM interest and persistence. Given our initial intuition from Table 4, we then include the math confidence proxy in Equation 1.2.

$$1.2\ Pr(Persist\ in\ STEM) = \Phi(\beta_1 University\ Selectivity + \beta_2 Female + \beta_3 Minority + \beta_4 Socioeconomic\ Composite + \beta_5 Math\ Score + \beta_6 Math\ Confidence)$$

Figure 1: Baseline Regression and Math Confidence Results

	(1)	(2)
Variable	Baseline	Math Confidence
Selectivity	-0.020 (0.069)	-0.032 (0.070)
Female	-0.191*** (0.064)	-0.135** (0.065)
Minority	0.086 (0.077)	0.064 (0.079)
SES Composite Score	-0.064 (0.044)	-0.071 (0.044)
Math Score	0.028*** (0.004)	0.024*** (0.004)
Math Confidence		0.197*** (0.037)
Constant	-0.816*** (0.258)	-0.696*** (0.265)
Observations	2,048	2,001

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In the results for this regression, shown in column 2 of Figure 1, we see again that *Selectivity*, *Minority*, and *SES Composite* are not statistically significant, while *Female*, *Math Score*, and *Math Confidence* are significant. Math score and confidence are relevant predictors at all levels of significance, again lending credence to our intuition that both preparedness and confidence are important players in STEM persistence. Our intuition at this stage is that students' actual abilities might not even be as important as their beliefs in those abilities. Prior literature informs that hypothesis by demonstrating the existence of the "stereotype threat" (Shapiro and Williams, 2012), a phenomenon in which minority groups such as African Americans and women are discouraged from pursuing STEM fields and achieving positive education and career outcomes due to negative stereotypes about these groups' abilities. This feeling that one does not belong can be transmuted from parents, teachers, and peers, and is a sentiment that could surely translate to lower math self-confidence for a given level of preparedness.

2. Confidence Relative to Ability and its Impact on the Persistence Gap

Having demonstrated the impact of students' beliefs in their abilities on their persistence in STEM, we move forward by delving deeper into our confidence measures. Progressing from simply utilizing absolute confidence via our proxies in the dataset, we create our own measure for *Overconfidence*, a proxy for confidence relative to actual performance. Specifically, we regress *Math Confidence* – or students' absolute levels of math self-efficacy – against their math scores, and then utilize the residuals to estimate students' levels of overconfidence (or under confidence) relative to their math scores. The idea behind our overconfidence measure is that a given math score should theoretically coincide with some absolute level of math confidence. In other words, the higher a student scores on the standardized math assessment, the more confident that student should be in his or her math abilities. However, students' true perceptions of their skills often deviate from that assumption, and the magnitude of that internal confidence *miscalculation* may ultimately be a better predictor of STEM persistence than a more absolute version of math self-confidence. For robustness, we also include square and cubic terms for *Math Score* in the creation of our overconfidence metric.

In table 5, we see that female students are under-confident on average, while male students are over-confident on average. At more selective schools, despite having similar math scores as their male counterparts, a relative lack of confidence for female students seems to coincide with lower persistence in STEM fields. Interestingly, under-confidence does not appear to impact female persistence in STEM fields relative to their male counterparts at less selective universities.

Table 5: Selectivity and Gender against Overconfidence and Persistence

Selectivity and Gender		Math Score	Overconfidence	Persisted in STEM
More Selective Universities	Female	60.93	-0.07	0.758
	Male	63.06	0.19	0.841
Less Selective Universities	Female	54.48	-0.12	0.746
	Male	55.32	0.16	0.766

Source: High School Longitudinal Study of 2009

2.1 Baseline Regression with Overconfidence Measure

Our first regression in this section is identical to our baseline regression, with *Overconfidence* replacing absolute *Math Confidence*. We expected to have extremely similar results to Equation 1.2, with the caveat that creating the overconfidence measure allows us to disentangle what could be considered irrational over/under-confidence relative to abilities. In the results shown below in column 1 of Figure 2, we see that our hypothesis holds true in that the coefficient on *Overconfidence* is both significant and positive, suggesting that being overconfident relative to one's abilities appears to positively impact persistence in STEM fields.

$$1.3\ Pr(Persist\ in\ STEM) = \Phi(\beta_1 University\ Selectivity + \beta_2 Female + \beta_3 Minority + \beta_4 Socioeconomic\ Composite + \beta_5 Math\ Score + \beta_6 OverConfidence)$$

2.2 Female and Minority Groups with Interaction Terms

Based on our observations that *Math Score*, *Overconfidence*, and *Female* are significant in relation to STEM persistence, we add interaction terms to ascertain why women choose to leave their STEM majors at higher rates, especially at more selective universities. In Equation 1.5, we introduce three interaction terms: *Selectivity* * *Math Score*, *Selectivity* * *Overconfidence*, and *Overconfidence* * *Female*. The basis for the *Selectivity* * *Math Score* interaction term is to identify whether at selective schools, pre-college credentials are even more important. Similarly, for *Selectivity* * *Overconfidence*, we seek to determine whether overconfidence is more predictive of persistence at more selective universities. Lastly, the interaction between *Overconfidence* and *Female* will determine whether overconfidence matters more for females.

$$1.4\ Pr(Persist\ in\ STEM) = \Phi(\beta_1 University\ Selectivity + \beta_2 Female + \beta_3 Minority + \beta_4 Socioeconomic\ Composite + \beta_5 Math\ Score + \beta_6 OverConfidence + \beta_7 Selectivity * Score + \beta_8 Selectivity * Over\ Confidence + \beta_9 OverConfidence * Female)$$

Figure 2: Overconfidence and Interaction Term Data

	(1)	(2)
Variable	Overconfidence	Interaction Terms
Selectivity	-0.031 (0.070)	-1.442*** (0.536)
Female	-0.137** (0.065)	-0.133** (0.066)
Minority	0.064 (0.079)	0.060 (0.079)
SES Composite Score	-0.071 (0.044)	-0.073* (0.045)
Math Score	0.031*** (0.004)	0.020*** (0.006)
Overconfidence	0.192*** (0.037)	0.134** (0.063)
Selectivity * Score		0.023*** (0.009)
Selectivity * Overconfidence		0.103 (0.074)
Overconfidence * Female		0.020 (0.074)
Math Confidence		
Constant	-1.045*** (0.267)	-0.406 (0.360)
Observations	2,001	2,001

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Unsurprisingly, the results in column 2 of Figure 2 show that the coefficients on *Female, Math Score*, and *Overconfidence* are all significant, with the same signs as in prior regressions. However, unlike in previous regressions, the selectivity of the university now emerges as a significant, negative predictor of STEM persistence. Bringing in the additional result here that the first interaction term, *Selectivity * Math Score*, is a significant positive predictor, we have demonstrated not only that selectivity of school matters – with more selective schools negatively impacting persistence – but also that that preparedness tends to matter even more at selective schools. This finding affirms our initial hypothesis that mismatch theory plays a significant role in driving persistence across minority and gender groups. More specifically, it tells us that students who perform below some threshold math score are more likely to persist in STEM at less selective universities, whereas students who perform above that threshold are more likely to persist in STEM at more selective universities. For the students represented in the HSLS specifically, by looking at the coefficients on *Selectivity* and *Selectivity * Math Score* (and assuming average overconfidence) we can pinpoint that threshold math score at 63, which is more than one standard deviation *above* the mean math score for the sample in the dataset.

While it's certainly interesting to note a threshold score for the *average* student in the sample, mismatch theory is a *marginal* phenomenon. In other words, to determine whether female and minority students are being mismatched more often than their male and non-minority counterparts, we have to decompose that threshold and investigate beyond the average. To that end, the greatest drawback of mismatch theory research is the impossibility of observing outcomes for mismatched (underprepared) students if they had attended less selective universities. In the analysis that follows, we attempt to simulate this counterfactual scenario by using Equation 1.4 to predict the changes in likelihood of STEM persistence for students if they attended less selective schools instead of more selective schools.

3. A Deeper Look into Mismatch Theory for Women, Through the Lens of Overconfidence

In order to execute our simulation, we first take all students who intended to pursue STEM degrees at more selective universities. It is important to note that for all the results that follow in this section, we only look at students who actually attended more selective universities, since the most pertinent mismatch cases for us are the students at more selective schools who may have been more likely to persist in STEM at less selective schools. It would also be interesting to consider the other subset of mismatch cases, since some students who attended less selective universities may have also been more likely to persist in STEM had they attended more selective universities. However, that second layer of mismatch is not as relevant to the underrepresentation of women and minority individuals in STEM and is therefore beyond the scope of our paper.

Next, for each student, we plug all observed values into equation 1.4 in order to evaluate the *true* odds of that student persisting in STEM at the more selective school that s/he decided to attend. Then, to gauge the counterfactual outcome, we run equation 1.4 again for each student, this time changing the value of *Selectivity* to zero in order to evaluate the *predicted* odds of that same student persisting in STEM had s/he instead attended a less selective university. If a given student's likelihood of STEM persistence was higher at the more selective university s/he attended in reality, we concluded that s/he was properly matched at that more selective school. On the other hand, if a student's likelihood of STEM persistence would have been higher at a less selective university, we concluded that s/he was mismatched at the more selective school s/he attended. It is important to note that we are only considering STEM persistence when making these mismatch judgments. Even for students who enter college with a preference for STEM, the importance of completing that STEM degree may realistically pale in comparison with the importance of the all-around educational experience. As a result, we understand that some students might not be mismatched even if they would have been more likely to persist in STEM at less selective schools. However, since the HSLS does not explicitly address priorities in that context, we move forward with our analysis as if STEM persistence is the primary objective for all students who initially intended to pursue a STEM degree.

Table 5 illustrates the results from the simulation process outlined above. The *mismatch rate* refers to the proportion of a given group that would have been more likely to persist in STEM at a less selective university. For instance, 41% of female students at more selective schools would have been more likely to complete their intended STEM degrees at less selective schools. The mismatch rate drops to 27% for male students, and the 14 percentage-point difference is statistically significant at the 5% confidence level.

Table 5: The Proportion of Female and Minority Students "Mismatched" at More Selective Universities, Relative to Their Male and Non-Minority Counterparts

	Female	Male	Minority	Non-Minority
Mismatch Rate (%)	0.41	0.27	0.38	0.32
Difference	*0.14**		0.06	
Average Math Score	64.00	66.04	62.89	65.65
Difference	*-2.04**		*-2.76**	
Average Overconfidence	0.11	0.34	0.39	0.21
Difference	*-0.23**		*0.18**	

Source: High School Longitudinal Study of 2009

*Significant at the 5% level

On the other hand, there is no significant difference in the mismatch rate between minority and non-minority students. This result contradicts prior mismatch theory literature, which typically focuses on race rather than gender and finds significantly higher mismatch rates for minority students. The lack of significance should not come as much of a surprise given our prior regression results, though, since *Minority* repeatedly failed to appear as a significant predictor of STEM persistence.

There is one surprising takeaway from Table 5, however. All prior mismatch theory literature points to low pre-enrollment preparedness levels – or low math scores – as the driving force behind mismatch at more selective universities. But, even though the difference in average math score is significant and negative for both gender and race, the difference in mismatch rate is only significant along gender lines. Perhaps, then, as illustrated in the last two rows of the table, the mismatch of female students actually stems from some combination of low pre-enrollment preparedness *and* underconfidence relative to those preparedness levels. Tables 6 and 7 take a more detailed look at mismatch rates for men and women by adding a slightly more granular confidence dimension.

Table 6: Gender Mismatching as a Function of Over/under-confidence

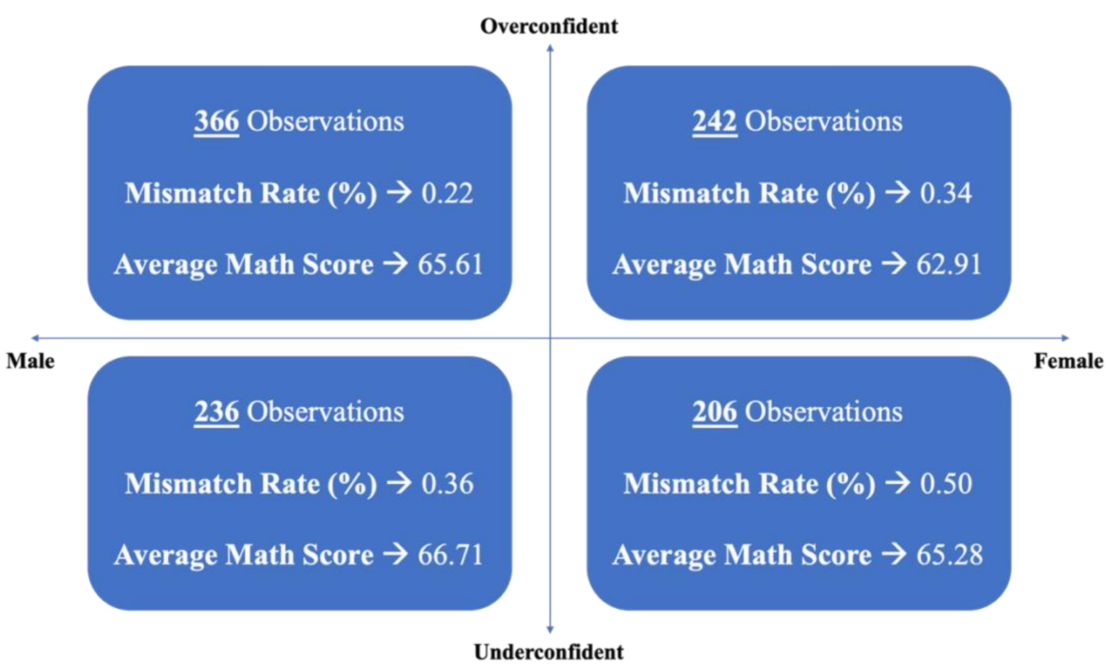

Table 7: Gender Mismatching as a Function of Over/under-confidence

Gender and Confidence		Mismatch Rate (%)	*Difference*	Average Math Score	*Difference*
Female	Overconfident	0.34	*-0.16**	62.91	*-2.37**
	Underconfident	0.50		65.28	
Male	Overconfident	0.22	*-0.14**	65.61	*-1.10**
	Underconfident	0.36		66.71	

Source: High School Longitudinal Study of 2009

*Significant at the 5% level

Here, we see that 50% of underconfident women would have been more likely to persist in STEM had they attended less selective universities, a number that drops considerably to 34% for women who are overconfident relative to their math scores. The mismatch rate for underconfident women is significantly higher *despite* significantly higher average math scores. However, the same general effect seems to be true for men as well, so over/under-confidence is not necessarily any more detrimental to women than men. And, as alarming as the difference in mismatch rates may seem between men and women within the same confidence bracket, the average math scores are significantly higher for men in each bracket. As a result, any gap in gender mismatch figures to stem largely from that difference in pre-enrollment preparedness. But, even though the higher mismatch rates for women may seem justified based on their lower math scores, our overconfidence measure is completely separate from those scores. So, there should be no reason for the men in this sample to be overconfident more often than women, even if they have higher math scores. Table 8, then, shows the root of the problem:

Table 8: The Proportion of Overconfident Men and Women at More Selective Schools

Gender and Confidence		Sample Observations	*% Overconfident*	*Difference*
Female	Overconfident	242	*0.54*	*-0.07**
	Underconfident	206		
Male	Overconfident	366	*0.61*	
	Underconfident	236		

Source: High School Longitudinal Study of 2009

The proportion of overconfident women is significantly lower than the proportion of overconfident men at more selective schools. From Table 7, we have already seen that overconfidence reduces the risk of mismatching by 16% for women, and 14% for men. If men and women were equally likely to be overconfident relative to their abilities, then any differences in mismatch rates should stem from differences in preparedness. However, from Table 8, we can now conclude that there are disproportionately more underconfident women than men in the sample, which consequently makes it more difficult on average for female students to persist in STEM at more selective universities.

In short, women tend to show higher mismatch rates than men *not* because overconfidence matters more for women than men, but rather because women at more selective schools are more frequently underconfident than their male counterparts. Now, the policy implication is not that underconfident women should attend less selective schools, but that more selective schools should modify their environments to build the confidence of these female students. Additionally, we can also explore the factors that drive overconfidence in the first place, in hopes of boosting that confidence in female students before they even reach college.

4. The Environmental and Attitudinal Factors Driving Overconfidence

The primary controls in the following regressions are *Female, Minority, SES Composite,* and a work ethic dummy that indicates whether or not a student spends more than one hour per week on math homework. We then add more variables of interest that capture environmental factors that may influence overconfidence across female and minority groups.

4.1 Parental Attitudes and Beliefs

The first environmental factor we investigate is that some parents may be biased against the idea of women in STEM. The HSLS survey asked all parents whether they believe female students are better or worse than male students at math. Our intuition here is that parents who are intrinsically biased against the math abilities of female students might create an adverse home environment that stifles girls' math self-confidence levels. In Equation 1.5 below, we include both the dummy term for biased parents as well as an interaction between *Parent Biased* and *Female.*

$$1.5\ OverConfidence = \beta_0 + \beta_1 Female + \beta_2 Minority + \beta_3 Socioeconomic\ Composite + \beta_4 Work\ Ethic + \beta_5 Parent\ Biased + \beta_6 Parent\ Biased * Female$$

The results shown below in column 1 of Figure 3 are noteworthy for several reasons. First, we see that *Female* is significant with a negative sign, suggesting that female students tend to have lower overconfidence levels relative to their male peers. Recall that in Equation 1.4, we saw that the interaction term between *Overconfidence* and *Female* was not a significant predictor of STEM persistence. This can now be explained by the notion that women tend to just have much lower confidence and overconfidence measures than men in the first place. Next, the results also illustrate that biased parents are not necessarily "worse" parents than non-biased parents, but that biased parents are particularly detrimental to the overconfidence of women as the coefficient on the interaction term between *Parent Biased* and *Female* is significant and negative. Additionally, we see that the coefficient on work ethic as a control is significant, suggesting that overconfidence is driven partially by effort and investment into education. One interesting point to note here is that while the coefficient on race is significant, its sign is positive, countering our original hypothesis. One potential cause for this result is that we saw in Tables 3 and 5 that on average, minority students actually tend to have higher overconfidence measures than their non-minority counterparts. The reason why minority individuals tend to be more overconfident calls for further investigation, but the surprising result may be explained by minority students' tendency to, on average, have much lower *SES Composite* scores than non-minority students in the dataset. In turn, these students attend public schools of lower quality, at which their confidence levels may be driven by their scores relative to a lower-performing set of peers.

The next home-related environmental factor we investigate is parental expectations for their children's education outcomes. We created a dummy variable that allows us to identify whether the parent responding to the HSLS survey believed their child would achieve a bachelor's degree or higher, as opposed to the alternatives of dropping out of college, earning an associate degree, earning a GED, or dropping out of high school. We again added an interaction with *Female* to determine whether parental expectations impact women and men differently.

$$1.6\ OverConfidence = \beta_0 + \beta_1 Female + \beta_2 Minority + \beta_3 Socioeconomic\ Composite + \beta_4 Work\ Ethic + \beta_5 Parent\ Biased + \beta_6 Parent\ Biased * Female + \beta_7 Parent\ Expectations + \beta_8 Parent\ Expectations * Female$$

Again, similar results hold in column 2 of Figure 3 for the variables carried over from Equation 1.5, as biased parents continue to have a significant negative effect on their daughters. For our variable of interest, we see that parental expectations alone are significant drivers of overconfidence, but that no significant difference exists between the impact of those expectations on men and women.

The next attitudinal factor we investigate is the student's own bias between males and females in math. Again, we include an interaction term in Equation 1.7 between *Student Biased* and *Female* to understand whether bias against females is marginally more detrimental to the overconfidence of women in STEM.

$$1.7\ OverConfidence = \beta_0 + \beta_1 Female + \beta_2 Minority + \beta_3 Socioeconomic\ Composite + \beta_4 Work\ Ethic + \beta_5 Parent\ Biased + \beta_6 Parent\ Biased * Female + \beta_7 Parent\ Expectations + \beta_8 Student\ Biased + \beta_9 Student\ Biased * Female$$

In column 3 of Figure 3, we see that females are challenged by attitudes held within their own household. In this case, females' biases against their own abilities in math and science serve as a major detriment to their confidence levels. The *Parent Biased* * *Female* interaction drops out as a significant predictor at the 5% level but remains significant at the 10% level, which is noteworthy in that parental biases have a negative effect on female confidence levels over and above a female child's own biases.

Lastly, in this section we also explore the role of students' expectations for themselves in relation to their level of overconfidence. We hypothesized that students' personal expectations would play a significant role in their overconfidence because students' own perceptions often arise from the intersection of various environmental factors. The specific factors that coalesce into students' self-perceptions include those that we investigated earlier in the paper, such as parental biases and expectations, as well as factors outside a student's home such as teacher/peer biases and expectations. Again, we interact the variable for students' expectations with gender. The results in column 4 below replicate the results from models 1-3 for the control variables. In terms of our variables of interest, we see again that *Parent Biased * Female* and *Parent Expectations* are significant. We also see that all the student expectation variables are significant. The significance on the interaction terms suggests that students' biases against females in math and students' perceptions of themselves are especially important for female students. This elucidates the stereotype threat, in that females who feel they do not belong in STEM due to peer, parental, or teacher biases are discouraged from persisting and succeeding in STEM. Additionally, the notion that students' personal expectations are even more important for women than men suggests that female self-perceptions should be a critical focus from an early age.

Figure 3: Parent and Student Attitudes and Expectations Results

Variable	(1) Parental Bias	(2) Parental Expectations	(3) Student Bias	(4) Student Expecta
Female	-0.206*** (0.021)	-0.174*** (0.042)	-0.165*** (0.024)	-0.094* (0.051)
Minority	0.154*** (0.019)	0.154*** (0.021)	0.154*** (0.021)	0.169*** (0.023)
SES Composite Score	0.005 (0.011)	-0.008 (0.012)	-0.007 (0.012)	-0.021 (0.014)
Work Ethic	0.060*** (0.018)	0.049*** (0.019)	0.051*** (0.019)	0.040* (0.021)
Parent Biased	0.012 (0.019)	0.004 (0.020)	-0.001 (0.021)	-0.004 (0.022)
Parent Biased * Female	-0.083** (0.036)	-0.072* (0.037)	-0.059 (0.038)	-0.066* (0.040)
Parent Expectations		0.135*** (0.032)	0.111*** (0.024)	0.099*** (0.028)
Parent Expectations * Female		-0.047 (0.046)		
Student Biased			0.107*** (0.031)	0.078** (0.034)
Student Biased * Female			-0.267*** (0.048)	-0.231*** (0.052)
Student Expectations				0.194*** (0.039)
Student Expectations * Female				-0.119** (0.055)
Constant	0.008 (0.047)	-0.068 (0.054)	-0.068 (0.052)	-0.180*** (0.058)
Observations	11,244	10,096	10,005	8,108
R-squared	0.021	0.023	0.026	0.032

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.

4.1 Role of the Mother

Moving forward from bias, we then progress to the involvement of mothers in students' academic lives to determine whether significant involvement from the mother can be beneficial for confidence levels, particularly for female students. To achieve that end, we introduce two new variables – a dummy for whether the student speaks to his or her mother about high school math classes, and a dummy for whether the student speaks to his or her mother about attending college in the future – in equations 1.9 and 1.10, as well as interactions of the variables with *Female*.

$$1.8\ OverConfidence = \beta_0 + \beta_1 Female + \beta_2 Minority + \beta_3 Socioeconomic\ Composite + \beta_4 Work\ Ethic + \beta_5 Parent\ Biased + \beta_6 Parent\ Biased * Female + \beta_7 Parent\ Expectations + \beta_8 Mom\ Math\ Talk + \beta_9 Mom\ Math\ Talk * Female$$

$$1.9\ OverConfidence = \beta_0 + \beta_1 Female + \beta_2 Minority + \beta_3 Socioeconomic\ Composite + \beta_4 Work\ Ethic + \beta_5 Parent\ Biased + \beta_6 Parent\ Biased * Female + \beta_7 Parent\ Expectations + \beta_8 Mom\ College\ Talk + \beta_9 Mom\ College\ Talk * Female$$

In both the results for Equations 1.9 and 1.10 below, we see very similar results. The various modes of maternal engagement in education do not affect women more than men, as evidenced by the insignificance of the coefficients on the interaction terms. However, we do see that both of the new variables are significant individually, suggesting that maternal involvement is beneficial to all students' overconfidence levels.

Figure 4: Mother's Role Results

Variable	(1) Mother and Math	(2) Mother and College
Female	-0.213*** (0.032)	-0.215*** (0.047)
Minority	0.157*** (0.021)	0.152*** (0.021)
SES Composite Score	-0.016 (0.012)	-0.013 (0.012)
Work Ethic	0.041** (0.019)	0.041** (0.019)
Parent Biased	0.004 (0.021)	0.005 (0.021)
Parent Biased * Female	-0.068* (0.038)	-0.069* (0.038)
Parent Expectations	0.112*** (0.024)	0.105*** (0.024)
Mother Talks Math	0.078*** (0.028)	
Mother Talks Math * Female	-0.003 (0.039)	
Mother Talks College		0.116*** (0.033)
Mother Talks College * Female		-0.008 (0.050)
Constant	-0.093* (0.054)	-0.134** (0.057)
Observations	9,952	9,951
R-squared	0.024	0.025

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.

Conclusion

As outlined at the beginning of the paper, our study aims to explain the factors responsible for the underrepresentation of women and minority individuals in STEM fields, in an attempt to guide policymakers as they take strides to reinforce the STEM workforce and equalize STEM representation. Since the most direct precursor to a STEM career is graduating from college with a STEM degree, our study focuses on STEM persistence rates at the university level.

Our results support mismatch theory in two ways: (1) all else equal, attending a more selective university decreases the likelihood that students will persist in their intended STEM majors; and (2) the negative marginal effect of attending a more selective school is reduced for students with higher math scores, and actually becomes positive for students whose scores lie more than one standard deviation above the mean. Since the female and minority students who attend more selective schools have lower math scores on average, a greater proportion of those students will naturally be mismatched as it relates to the ideal conditions for STEM persistence.

Prior literature on mismatch theory has revealed that for minority students, those lower math scores – which represent a lower level of pre-enrollment preparedness – drive the higher mismatch rates we tend to see among minority student samples. However, for female students, section 3 of our results illustrates that lower preparedness levels may not be the only – or even the most influential – driver of lower persistence rates and higher mismatch rates. For instance, as shown in Table 8, women who attend more selective schools are significantly more likely to be underconfident than their male counterparts, and that lack of self-efficacy significantly decreases their likelihood of STEM persistence. In other words, for women, mismatch theory is actually mediated by SCCT in that negative self-perceptions can cause female students to be mismatched at more selective schools even if they are sufficiently prepared. Consequently, the optimal resolution to higher mismatch rates for women has little to do with the university admission policies that separate students across more and less selective schools. Rather, in order to reduce attrition rates, the more selective schools simply need to take strides to support the confidence levels of women in STEM fields.

Considering that overconfidence in itself is a key driver of STEM persistence, the other path to lower attrition rates centers around the advancement of self-efficacy in underconfident individuals long before they even reach college. The development of math overconfidence seems to begin at home, with parent biases against females in STEM significantly stifling the confidence of female children before they even reach high school. Student biases against females in STEM significantly diminish the confidence of female students as well, supporting the idea of a stereotype threat and the alleged "masculinity" of STEM fields.

In terms of data analysis, one of the biggest limitations of our study is that the HSLS will not actually be complete until 2025, when the cohort of students finally settles into the workforce. As a result, we were only able to work with STEM persistence data at the university level, as opposed to the actual STEM career data that the HSLS will eventually offer. Additionally, the original appeal of the dataset came mostly from its national scope and large sample size, but the isolation of students interested in STEM reduced our sample from 23,503 to approximately 2,000.

As for limitations in our findings, the positive coefficients on *Minority* in section 4 of our results prevent us from making any conclusive statements about the racial persistence gap, which was one of the primary goals of our study. The result is interesting, however, and hopefully future research will seek to uncover the reasons why minority students have a heightened sense of overconfidence in their math abilities.

References

Anderson, E., and Kim, D. (2006). Increasing the success of minority students in science and technology. Washington: American Council on Education.

Arcidiacono, P., Aucejo, E., Coate, P., & Hotz, V. J. (2014). Affirmative Action and University Fit: Evidence from Proposition 209. doi: 10.3386/w18523

Arcidiacono, P., Aucejo, E. M., Fang, H., & Spenner, K. I. (2011). Does affirmative action lead to mismatch? A new test and evidence. *Quantitative Economics*, *2*(3), 303–333. doi: 10.3982/qe83

Arcidiacono, P., Aucejo, E. M., & Hotz, V. J. (2013). University Differences in the Graduation of Minorities in STEM Fields: Evidence from California. doi: 10.3386/w18799

Arcidiacono, P., Aucejo, E. M., & Spenner, K. (2012). What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice. *IZA Journal of Labor Economics*, *1*(1). doi: 10.1186/2193-8997-1-5

Astorne-Figari, C., & Speer, J. D. (2019). Are changes of major major changes? The roles of grades, gender, and preferences in college major switching. *Economics of Education Review*, *70*, 75–93. doi: 10.1016/j.econedurev.2019.03.005

Beede, D. N., Julian, T. A., Langdon, D., Mckittrick, G., Khan, B., & Doms, M. E. (2011). Women in STEM: A Gender Gap to Innovation. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1964782

Bleske-Rechek, A., & Browne, K. (2014). Trends in GRE scores and graduate enrollments by gender and ethnicity. *Intelligence*, *46*, 25–34. doi: 10.1016/j.intell.2014.05.005

Card, D., & Payne, A. A. (2017). High School Choices and the Gender Gap in STEM. NBER Working Paper Series. doi: 10.3386/w23769

Carnevale, A.P.; Smith, N.; and Melton, M. (2011). STEM. Report from Georgetown Center for Education and the Workforce.

Castleman, B. L., Long, B. T., & Mabel, Z. (2017). Can Financial Aid Help to Address the Growing Need for STEM Education? The Effects of Need-Based Grants on the Completion of Science, Technology, Engineering, and Math Courses and Degrees. *Journal of Policy Analysis and Management*, *37*(1), 136–166. doi: 10.1002/pam.22039

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion. *Journal of Policy Analysis and Management*, *32*(4), 692–717. doi: 10.1002/pam.21715

Fayer, S., Lacey, A., & Watson, A. (2017). STEM occupations: Past, present, and future. Spot-light on statistics. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.

Fouad, N. A., & Santana, M. C. (2016). SCCT and Underrepresented Populations in STEM Fields. *Journal of Career Assessment*, *25*(1), 24–39. doi: 10.1177/1069072716658324

Fryer, R. G., & Levitt, S. D. (2006). The Black-White Test Score Gap Through Third Grade. *American Law and Economics Review*, *8*(2), 249–281. doi: 10.1093/aler/ahl003

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. doi: 10.1016/j.econedurev.2010.06.010

Hanushek, E., & Rivkin, S. (2008). Harming the Best: How Schools Affect the Black-White Achievement Gap. doi: 10.3386/w14211

Hastings, J., Neilson, C., & Zimmerman, S. (2013). Are Some Degrees Worth More than Others?

Evidence from college admission cutoffs in Chile. doi: 10.3386/w19241

Hoxby, C., & Avery, C. (2014). The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students. doi: 10.3386/w18586

Kanny, M. A., Sax, L. J., & Riggers-Piehl, T. A. (2014). Investigating Forty Years Of Stem Research: How Explanations For The Gender Gap Have Evolved Over Time. *Journal of Women and Minorities in Science and Engineering*, *20*(2), 127–148. doi: 10.1615/jwomenminorscieneng.2014007246

Kirkeboen, L., Leuven, E., & Mogstad, M. (2016). Field of Study, Earnings, and Self-Selection. *Quarterly Journal of Economics*, *131*(3), 1057-1111. doi:10.1093/qje/qjw019

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a Unifying Social Cognitive Theory of Career and Academic Interest, Choice, and Performance. *Journal of Vocational Behavior*, *45*(1), 79–122. doi: 10.1006/jvbe.1994.1027

Liben, L. S., & Coyle, E. F. (2014). Developmental Interventions to Address the STEM Gender Gap: Exploring Intended and Unintended Consequences. The Role of Gender in Educational Contexts and Outcomes Advances in Child Development and Behavior, 77–115. doi: 10.1016/bs.acdb.2014.06.001

Luppino, M., & Sander, R. H. (2012). College Major Competitiveness and Attrition from the Sciences. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2167961

Makarova, E., Aeschlimann, B., & Herzog, W. (2019). The Gender Gap in STEM Fields: The Impact of the Gender Stereotype of Math and Science on Secondary Students Career Aspirations. *Frontiers in Education*, *4*. doi: 10.3389/feduc.2019.00060

Mann, A., & Diprete, T. A. (2011). Trends in Gender Segregation in the Choice of Science and Engineering Majors. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1938328

Mitchell, J., Fuller, A., & Hackman, M. (2019, November 20). Which College Graduates Make the Most? Retrieved from https://www.wsj.com/articles/which-college-graduates-make-the-most-11574267424.

Noonan, R. (2017). STEM jobs: 2017 update. ESA issue brief #02–17. Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration, Office of the Chief Economist.

Stinebrickner, R., & Stinebrickner, T. (2013). A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. doi: 10.3386/w19165

Stinebrickner, T., & Stinebrickner, R. (2011). Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major. doi: 10.3386/w16869